# Optimisation of Pedotransfer Function Models for Soil Hydraulic Properties Using an Artificial Neural Network Ensemble Method

**Lee Baker**

**PhD**

**2005**

**University of Abertay Dundee**

**Optimisation of Pedotransfer Function Models for Soil Hydraulic Properties Using an Artificial Neural Network Ensemble Method**

A thesis submitted in partial fulfilment of the requirements
of the University of Abertay Dundee
for the degree of Doctor of Philosophy

by

Lee Baker

University of Abertay Dundee
Dundee, United Kingdom

7.11.2005

I certify that this thesis is the true and accurate version of the thesis approved by the examiners.

Signed [redacted] Date 9/11/2005
Director of Studies

# Abstract

Soil hydraulic properties are very time-consuming and expensive to measure directly. Conversely, routinely collected soil survey parameters (e.g. soil texture, dry bulk density and organic matter content) are relatively cheap and easy to collect. Fortunately, mathematical regression models called Pedotransfer Functions (PTFs) allow the transfer of data we *have* (soil survey parameters) into the data we *need* (soil hydraulic properties), and this thesis focusses on the potential and modelling methodologies of Artificial Neural Network (ANN) ensembles as a means by which to construct PTFs. The individual processes of the ensemble modelling procedure are examined herein, and suggestions as to how they may be improved upon, by mathematical, conjectural and empirical means, are discussed. Ultimately, it will show that more accurate, robust and efficient PTFs may be constructed by incorporating these into such models.

An ensemble is a collection of individual ANNs, each of which provides an independent solution to the same problem. Thus, the combined strengths of the individual models are augmented, whilst their weaknesses are diminished. For any modelling methodology, a trade-off exists between the bias and variance components of the model error – decreasing the bias results in higher variance, and *vice versa.* However, for the ensemble method, combining ensemble members results in the reduction or elimination of the variance, whilst leaving the bias unaltered. Thus, the aim of the ensemble modeller is to determine an optimum balance between achieving low bias (regardless of the consequences to the variance) and conserving data – a trade-off between the *amount* of data used and the *value* of that data. It has been empirically demonstrated here that by training ANNs using data selected such that all parts of the dataspace are equally represented, bias in ANN-PTF ensembles may be reduced to negligible levels. In addition, results showed that, using one-third or less the amount of data, the ensemble method yields results that are at least as accurate as single ANN methods.

## Acknowledgements

During the past three years, my work has benefited greatly from discussions with Attila Nemes, Yakov Pachepsky and James Bown.

I especially thanks my supervisors, Iain M Young, Dave Ellison and Bill Samson for their support, criticism, help and advice, and John W Palfreyman for his continued faith and support.

I would like to thank The Carnegie Trust for the Universities of Scotland and the Department of Contemporary Sciences, University of Abertay Dundee for funding my trip to the United States Department of Agriculture, Beltsville, Washington DC where I learnt some invaluable lessons about my research topic.

I would also like to thank Henk Wösten, Allan Lilly and Walter Rawls for agreeing to allow me to conduct research on the HYPRES and USDA-NRCS databases.

Finally, I would like to thank my mother and father for teaching me the beauty and value in life. Without the continuous support from my parents, I would have been lost long ago. Thankyou.

This work was supported by a University of Abertay Dundee studentship.

This thesis is dedicated to the memory of

Andrew James Hurst

and

Eric Staniforth Gill

A problem well posed contains its own solution

# Contents


**Abstract** iii

**Acknowledgments** iv

**0. Overview** 1

**1. Introduction** 6

**2. The HYPRES Database and Pedotransfer Functions** 13

Introduction 13

2.1 The HYPRES Database 13

2.2 Interactions Between Soil and Water 18
2.2.1 Water – Water Interactions 18
2.2.2 Soil – Water Interactions 21

2.3 Introduction To Pedotransfer Functions 29
2.3.1 Functional Descriptors of Water Retention 30
2.3.2 Functional Descriptors of Hydraulic Conductivity 34
2.3.3 Parameter Prediction of Hydraulic Characteristics 39
2.3.4 Prediction of Hydraulic Characteristics by Soil Structure Model 39
2.3.5 Point Prediction of Hydraulic Characteristics 40

2.4 Class Pedotransfer Functions 41

2.5 Continuous Pedotransfer Functions 42

2.6 Pedotransfer Function Evaluation Criteria 43

Chapter Conclusions 45

**3. Artificial Neural Networks** 46

Introduction 46

3.1 Primer on Neurons and Networks 46
3.1.1 Similarities Between Biological and Artificial Neurons 47

3.2 The Biological Neuron 49

3.3 The Artificial Neuron 50
3.3.1 Artificial Neuron Input Function 50
3.3.2 Artificial Neuron Output Function 51

3.4 Networks of Artificial Neurons 53
3.4.1 Network Architectures 53

3.5 Training an Artificial Neural Network 58
3.5.1 Training Regimes 58
3.5.2 Training Algorithms 60

3.6 Training Issues of Artificial Neural Networks 70
3.6.1 Data Standardisation 70
3.6.2 Generalisation 72
3.6.3 Noise 73
3.6.4 Underfitting and Overfitting 75
3.6.5 Early Stopping 75

3.7 Testing Artificial Neural Networks 76
3.7.1 Split-Sample Validation 76
3.7.2 Cross-Validation 76
3.7.3 Bootstrapping 77

Chapter Conclusions 77

---

**4. Combining Artificial Neural Networks 78**

Introduction 78

4.1 Artificial Neural Networks For Pedotransfer Functions 78

4.2 Artificial Neural Network Ensembles 79
4.2.1 The Bias/Variance Trade-Off 81
4.2.2 Creating Ensemble Members 83
4.2.3 Creating Ensembles (Combining Members) 85

4.3 Modular Artificial Neural Networks 86
4.3.1 Creating Modular Components 87
4.3.2 Combining Modular Components 88

Chapter Conclusions 88

---

**5. Selection and Statistical Analyses of the Soil Data 90**

Introduction 90

5.1 Data Selection and Selection Criteria 90

5.2 Results of Statistical Analyses of HYPRES 95
5.2.1 Do the Soil Textural Components Sum to 100? 96
5.2.2 Anderson-Darling Test for Normality 97
5.2.3 Minimum and Maximum 98
5.2.4 Mean and Median 99
5.2.5 Standard Deviation and Coefficient of Variation 102

Chapter Conclusions 104

**6. Continuous PTF of Soil Water Retention – an ANN Ensemble Method 105**

Introduction 105

6.1 Preliminary Investigations and Optimisation Procedures 105
6.1.1 Data Selection 105
6.1.2 Re-scaling Algorithm 107
6.1.3 Optimum Number of Test Sets 113
6.1.4 Optimum Number of Ensemble Members 114
6.1.5 Training Parameters and Algorithms 115
6.1.6 Evidence for the Optimum ANN Architecture 116
6.1.7 Evidence for the Optimum Number of Soil Horizons per ANN 118

6.2 Continuous PTF – Results and Discussion 121
6.2.1 Green and Durable – New Terms for ANN-PTF Ensembles 121
6.2.2 ANN Ensemble Soil Water Retention Continuous PTF Results 124
6.2.3 Single ANN Soil Water Retention Continuous PTF Results 133
6.2.4 Continuous PTF Results Comparison and Discussion 135
6.2.5 Sensitivity Analyses of the Ensemble 138

6.3 Adding Error Bars to Predictions 146
6.3.1 Error Bars – a Bootstrapping Method 146
6.3.2 Error Bars – an ANN Ensemble Method 148

6.4 Error Bar Modelling – Results and Discussion 150
6.4.1 Bootstrapping Error Bar Results 150
6.4.2 ANN Ensemble Error Bar Results 153
6.4.3 Error Bar Results Comparison and Discussion 156

Chapter Conclusions 164

**7. Class PTF of Soil Water Retention – an ANN Ensemble Method 166**

Introduction 166

7.1 Preliminary Investigations and Optimisation Procedures 166
7.1.1 Optimum Number of ANN Ensemble Members 166
7.1.2 Evidence for the Optimum Number of Soil Horizons per ANN 167

7.2 Class PTF – Results and Discussion 169
7.2.1 ANN Ensemble Soil Water Retention Class PTF Results 169
7.2.2 Sensitivity Analyses of the Ensemble 173

7.3 Error Bar Modelling – Results and Discussion 174

Chapter Conclusions 180

**8. Testing Models on Alternative Databases 181**

Introduction 181

8.1 Results of Statistical Analyses of USDA-NRCS Database 181
8.1.1 Do the Soil Textural Components Sum to 100? 183
8.1.2 Anderson-Darling Test for Normality 184
8.1.3 Minimum and Maximum 184
8.1.4 Mean and Median 185
8.1.5 Standard Deviation and Coefficient of Variation 187

8.2 USDA-NRCS Continuous PTF 188
8.2.1 Results and Discussion 188

8.3 USDA-NRCS Class PTF 196
8.3.1 Results and Discussion 196

8.4 Error Bar Modelling – Results and Discussion 202
8.4.1 Continuous PTF Error Bar Results 202
8.4.2 Class PTF Error Bar Results 207

Chapter Conclusions 212

**9. Model Comparisons 214**

Introduction 214

9.1 HYPRES Class and Continuous PTFs 214

9.2 USDA-NRCS Class and Continuous PTFs 217

9.3 HYPRES and USDA-NRCS Continuous PTFs 220

9.4 HYPRES and USDA-NRCS Class PTFs 223

9.5 HYPRES Class and Continuous Error Models 225

9.6 USDA-NRCS Class and Continuous Error Models 230

Chapter Conclusions 235

**10. ANN-PTF Ensembles of Hydraulic Conductivity 237**

Introduction 237

10.1 Data Selection 237

10.2 Modelling Procedure 238

10.3 Results 239

10.4 Discussion 241

Chapter Conclusions 243

**11. Discussion and Conclusions 244**

Introduction 244

11.1 Pedotransfer Functions 244

11.2 Artificial Neural Network Ensembles 245

11.3 The HYPRES Database 246

11.4 The Bias/Variance Trade-Off 247

11.5 Continuous HYPRES ANN-PTF Ensemble 248
11.5.1 Modelling Parameters 248
11.5.2 Model Evaluation Criteria 249
11.5.3 Discussion of Results 250
11.5.4 Sensitivity Analyses 250
11.5.5 Error Prediction Methods 250

11.6 Class HYPRES ANN-PTF Ensemble 252
11.6.1 Discussion of Results 252
11.6.2 Sensitivity Analyses 253
11.6.3 Error Prediction Methods 253

11.7 Continuous USDA-NRCS ANN-PTF Ensemble 253
11.7.1 Discussion of Results 253
11.7.2 Sensitivity Analyses 254
11.7.3 Error Prediction Methods 254

11.8 Class USDA-NRCS ANN-PTF Ensemble 255
11.8.1 Discussion of Results 255
11.8.2 Sensitivity Analyses 255

11.8.3 Error Prediction Methods 255

11.9 Model Comparisons 256
11.9.1 HYPRES Class and Continuous PTFs 256
11.9.2 USDA-NRCS Class and Continuous PTFs 257
11.9.3 HYPRES and USDA-NRCS Continuous PTFs 257
11.9.4 HYPRES and USDA-NRCS Class PTFs 258
11.9.5 HYPRES Class and Continuous Error Models 258
11.9.6 USDA-NRCS Class and Continuous Error Models 258

11.10 Modelling Hydraulic Conductivity 259

Chapter Conclusions 259

**12. Future Work** **263**

**References** **267**

# Chapter 0

# Overview

Man can survive for three days without food and water. He can survive for up to two months by ingesting water alone, but by ingesting both food and water he may live for a century. The food he eats is provided mainly by agriculture (crops, such as vegetables and grain) and livestock (meat and dairy products), and these use an extraordinary amount of water. It takes 1000 tons of water to raise one ton of grain, and an astonishing 16,000 tons of water to raise one ton of beef. It is fortunate then, that water is plentiful on the 'blue planet'. But is it? All life on earth depends on water, and it is estimated that approximately 1.1 billion people worldwide lack access to improved water sources. This lack of access comes with a heavy burden. Some 1.7 million deaths a year worldwide are attributable to unsafe water and to poor sanitation and hygiene, mainly through infectious diarrhoea. Most of the deaths (90%) occur in children, and virtually all occur in developing countries. Every year, over one million people die of malaria, a disease closely linked to the poor management of water resources, and about 6% of the global burden of disease is water related (WHO 2004).

Of all the water on earth, less than 3 percent is fresh, and all but three-thousandths of that is locked up in glaciers or icecaps (UNCSD 1997), or is at present deemed too deep in the earth to retrieve economically. Of the freshwater available in rivers, lakes, and accessible groundwater, an increasingly large fraction is polluted with biological, chemical, and radioactive contaminants (WHO 1995, UNCSD 1997).

Clearly, fresh, clean water is scarce and becoming more so.

Fortunately, humans and animals have the ability to relocate themselves to find the nearest clean water source, and, in the case of man, relocate the water to wherever he finds himself. In contrast with humans and animals, however, plants live permanently in one place, so they have to remove water from the soil in their immediate vicinity. Plant life depends essentially on water that is stored within the soil and is available for extraction. For extraction of water, plants rely on their root systems, which continue to grow throughout most of their lives. The quality of the soil

as a store of root-accessible water depends on texture and structure, and in turn, soil texture and structure determine the size of the pore spaces. The pore space is the conduit that stores and conducts water.

It follows then, that the interactions between water and soil should be characterised in some detail in order for the flux and storage of water at the land surface to be sustainably managed and utilised. Water retention is the term that characterises the storage of water, and hydraulic conductivity is the name given to the flow of water through the soil. Together they are known as the soil hydraulic properties. The soil hydraulic properties, therefore, play a crucial role in developing models that help us to understand the retention and transport of water and solutes. For example, the effect of the addition of contaminants to soil needs to be characterised in order to understand the impact on the water/soil system to identify viable methodologies of bioremediation[1]. Such models are inevitably data hungry, with demand for extensive and high quality datasets far outstripping supply (Elsenbeer 2001).

Over recent years there has been a concerted effort in developing databases of soil hydraulic properties (for example, see Wösten *et al.* 1999 to see how databases are constructed, and Wösten *et al.* 2001 for a review of how such databases may be used). However, soil hydraulic properties are very time-consuming and expensive to measure by direct means. Conversely, routinely collected soil survey parameters, such as soil texture, dry bulk density and organic matter content, are relatively cheap and easy to collect. The solution to this disparity is, in the words of Bouma (1989), to transfer the data we *have* (soil survey parameters) into the data we *need* (soil hydraulic properties), and mathematical regression models called pedotransfer functions have been developed to achieve this. Pedotransfer functions have been constructed by various means, such as Statistical Regression Analysis (e.g. see Wösten *et al.* 1998), Group Methods Of Data Handling (e.g. see Rawls *et al.* 2003), Classification And Regression Trees (e.g. see McKenzie and Jacquier 1997), all with varying degrees of success.

[1] Bioremediation is the elimination, attenuation or transformation of polluting or contaminating substances by the use of biological processes, to minimise the risk to human health and the environment.

The different types of pedotransfer functions have become important tools in research directed towards quantifying the most important physical and biological processes active in saturated and unsaturated soils, providing a measure of correspondence between measured and simulated functional soil behaviour. Latterly, Artificial Neural Networks have proved popular with many researchers, and have been shown to yield results that are at least as good as other techniques and overcome some of the statistical assumptions hard-wired into pedotransfer functions. Their recent popularity is, however, a little deceiving. Artificial neural networks are very data hungry, and these methods of building pedotransfer functions were not possible until the development of databases of soil hydraulic properties. Now that such databases are becoming readily available to environmental modellers, artificial neural network methods have become the tool of choice for many. However, for many modellers, it has become relatively easy to simply pass all or most of the available data to the artificial neural network with little regard to the precision, resolution and quality of the process. Little, if any, research has been conducted into answering the questions 'How much data is required?', 'How may we optimise the model?' and, more importantly, 'How may we improve the reliability and robustness of the model?'. On discussing various methods for constructing pedotransfer functions, Wösten *et al.* (2001) stated that '*the fact that these methods yield comparable results indicates that major progress has not to be expected from new statistical methods but rather from better data*'. This statement is rather complacent and perhaps ignores the importance of rigorous statistical procedures required for modelling.

The aims of this thesis are:

1. To introduce and provide a review of the current literature of pedotransfer functions, artificial neural networks, and ensembles and modules (groups of individual artificial neural networks that are combined in some way in order to improve the overall performance of the model).
2. Discuss how the modelling methodologies may be optimised for the data, and the data optimised for the modelling methodologies, particularly in that the process is iterative. In particular, the usage of artificial neural network ensembles for constructing both class and continuous pedotransfer functions is detailed. Additionally, working knowledge of two possible methods for determining errors

of individual predictions; an error ensemble method and a Bootstrapping method, will be gained.

3. To determine the optimum parameters (i.e. extracting the maximum information from minimum data) for class and continuous artificial neural network ensemble pedotransfer functions, such as the optimum number of soil horizons per artificial neural network, the optimum number of artificial neural network members per ensemble and the optimum number of tests to use.
4. To construct optimised class and continuous pedotransfer function models of soil water retention at field capacity by the artificial neural network ensemble method.
5. To construct the Bootstrapping and error ensemble models of error prediction, and implement them on both the class and continuous pedotransfer functions of soil water retention at field capacity.
6. To construct optimised class and continuous pedotransfer function models of soil hydraulic conductivity at field capacity by the artificial neural network ensemble method.
7. To demonstrate that, although large soil hydraulic databases are required for research purposes, since ensembles require significantly less data than do single artificial neural network methods to yield the same results, small soil hydraulic databases are sufficient for model construction using the ensemble method.
8. To enable environmental modellers and field scientists to work together more closely by highlighting the most important information required by modellers, and hence which data needs to be sampled with the greatest care by field scientists.

In addition to the stated aims of the research detailed in this thesis, there are a number of themes that permeate throughout, and are of fundamental importance to the concept of modelling by the ensemble method.

'Occam's razor' is the principle that states that unnecessarily complex models should not be preferred to simpler ones. In terms of artificial neural networks, Occam's razor is important when making design choices as to how many processing units to employ. A more complex and unconstrained artificial neural network (large number of processing units) will nearly always learn the examples in the training set better than simpler ones, however, the simpler artificial neural network may actually be a better model of the problem and generalise better to new examples.

Optimisation is the concept of extracting the maximum information from minimum data. Of course, the more data that is available to the modeller, the more information he possesses about the problem domain. However, one must consider the value of this data. If the amount of data acquired were to be doubled, does this necessarily mean that the amount of information about the problem domain is doubled? If so, then the increase in available data is valuable to the modeller. Now consider that doubling the amount of data results in an increase in the amount of information about the problem domain of 5%. Is this increase valuable to the modeller? As more is known about the problem domain, there remains less about it to be learnt, and thus, there is a trade-off between the amount of data acquired and the value of that data.

Recycling is a term that all environmental scientists are aware of, indeed I'm quite sure that most, if not all, are active in recycling everyday items, such as paper, glass, plastic, etc., but how many actually recycle their data? Modellers are taught that complete independence should be maintained between the data that is used to construct a model and the data that is used to test it. This has always been regarded as good practice and should be encouraged. However, if there was a way to use the same data in the modelling process and in the testing process (i.e. recycling the data), and still maintain (almost) complete independence between the modelling and test data, should we not utilise this method? In this thesis, this point will be addressed in terms of the ensemble method, and will show that indeed, data can be recycled.

# Chapter 1

# Introduction

During standard soil surveys, there are many basic soil parameters that are collected, such as soil texture, particle size analysis, bulk density and organic matter content. Other important soil properties, such as moisture retention (a measure of how much water a given soil sample holds) or hydraulic conductivity (a measure of the flux of water through a given soil sample), are much more expensive and time-consuming parameters to determine, and thus rarely measured. As an example of why these properties are important, consider the moisture retention of a given soil sample. Let a vertically oriented soil core be thoroughly wetted with a known amount of water, and measure the volume of water that collects in a vessel below the core. Now exert a small amount of pressure to the top of the soil core and re-measure the volume of water in the vessel. This process is repeated until the residual water has been expelled from the core and the pores begin to drain. This is known as the air entry point, or field capacity. Now continue the process until all the water has been drained from the pores, and the final vestiges of moisture can only be removed by heating the core. This is known as the permanent wilting point, so called because it is the point at which water molecules are held so tightly by the soil matrix that a plant cannot take up the water fast enough to sustain its life. Plotting the ratio of the difference between the total amount of water added to the soil core and the amount accumulated in the vessel to the total amount added, against the pressure exerted gives what is known as the water retention curve. The important points on this curve are the field capacity and permanent wilting points, and the difference between these is termed as the available water, since this is the amount of water that is available for plants to withdraw from the soil. This water retention curve is different for every soil sample, and thus, the available water is different for each type of soil. The ability of plants to flourish in the soil medium is highly dependent upon their ability to withdraw water from the soil, so it is important to be able to match a given plant with an appropriate soil type. In terms of the soil, this means having some process by which the available water can be

quickly and easily determined. Unfortunately, the process of directly determining the water retention involves a large amount of time and expensive equipment. Thus, it is necessary to find some way of indirectly determining the water retention.

Over the years there have been many attempts to derive mathematical equations and models that estimate water retention from easily and cheaply collected soil survey parameters, all with varying degrees of success. These models are referred to as pedotransfer functions, and they are often calculated with the aid of one of the many databases of soil hydraulic properties available to soil scientists. These databases often contain data such as the sand, silt and clay contents of given soil samples, along with their dry bulk density and organic matter contents, and moisture retention and hydraulic conductivities measured at different pressures. There are many different ways in which one may construct a pedotransfer function, one of the more popular methods being the use of artificial neural networks. Artificial neural networks have a distinct advantage over many other modelling methodologies in that they do not require an *a priori* mathematical model to be formulated, and hence are not influenced by mathematical or statistical assumptions that are necessary in other approaches. Indeed, previous researchers developing pedotransfer functions have concluded that when the number of input parameters (such as sand, silt, clay, etc.) is greater than three, artificial neural networks usually perform better than regression techniques.

Artificial neural networks may be thought of as being a so-called 'black box' technique of computation, in that a set of input:output ('question':'answer') pairs may be 'learned' by asking an artificial neural network to match the answers with the questions. The 'black box' that sits in-between the input and output stages is so-called because the mathematical computation that takes place there has no physical basis, rather the artificial neural network 'learns' from experience. In this case, an artificial neural network may be asked to match soil survey parameters (such as sand, silt, clay, etc.) with hydraulic parameters (for instance, water retention or hydraulic conductivity). In order to build up its experience, an artificial neural network often requires a large number of samples with which to learn from, and this is where the use of databases of soil hydraulic properties is essential. Typically, there are two stages in the artificial neural network modelling process – encoding and decoding. The former is the training stage, where the input:output pairs are presented and 'learnt', and the latter is the prediction phase, where only the inputs are presented (the outputs are withheld) and the artificial neural network is asked to, based on what it has learnt,

make predictions of what it thinks the outputs should be. The output predictions may be compared with the withheld outputs, and statistics based on this comparison give an indication as to how well the artificial neural network has learnt. It is precisely these statistics that allow researchers to be able to compare their artificial neural network model methodologies with others.

All models, whether requiring *a priori* information or not, are subject to errors, and all errors fall into one of three categories – bias, variance and noise. A measurement procedure is said to be biased if, on the average, it consistently favours some outcomes over others, giving an answer that differs from the expected value. The variance is a measure of the average distance between each of a set of data points and their mean value, and is essentially a measure of the spread of the data. Noise is represented by the uncertainty of the input and/or output values, for instance, measurement error. Noise is an entity that forms part of the data, and it is usually irreducible, however, bias and variance exist both as part of the data and of the modelling process, and in fact, there exists a trade-off between them such that efforts to decrease one will inevitably lead to an increase in the other. This seems intuitively correct, since, as the complexity of a model is increased (i.e. increasing the amount of data), the variety of the data samples in the model is increased, and so the bias towards individual data samples must decrease. Thus, there is a clear minimum level of error that may be obtained, imposed by the bias/variance trade-off. However, this trade-off refers to a *single* predictor. What happens when there are a number of individual predictors? The term 'ensemble' is used to characterise a grouping of individual models that are combined in some way, and provide a means of reducing overall prediction errors by *using* the bias/variance trade-off, rather than being *restricted* by it. Combining a number of uncorrelated individual predictors results in a reduction, if not complete elimination of the variance, whilst leaving the bias unaltered. Therefore, an effective approach is to create a set of individual models that each exhibit high variance, but low bias, since the variance component can be removed by combining. In other words, one should take more account of the available data to reduce bias, since the increase in variance will be removed by combining individual predictions.

The research presented in this thesis is essentially a fusion of two disciplines; pedotransfer functions and artificial neural network ensembles, and details how and why they may be combined to best effect. This thesis highlights the general

philosophy of the research and details its progression using a range of mathematical and statistical techniques. The process that was followed is outlined in the flow diagram of Figure 1a.

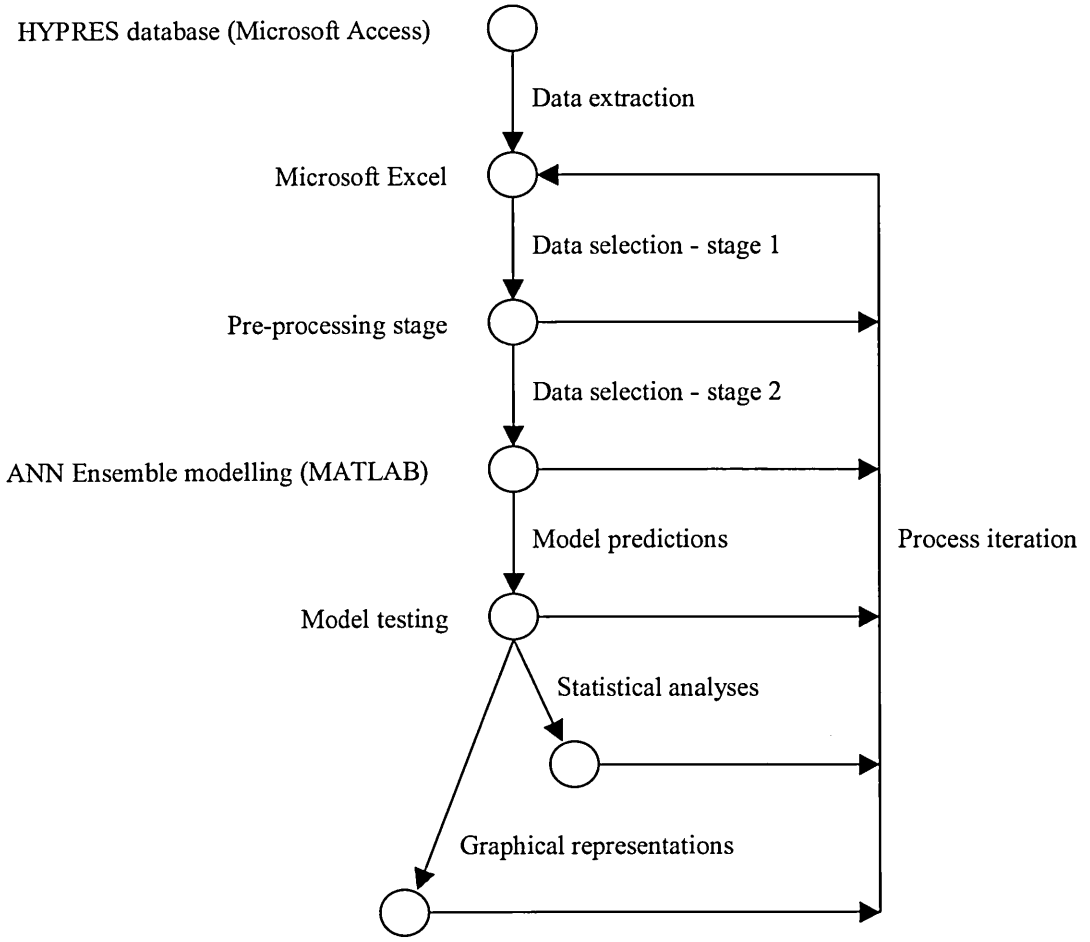



*Figure 1a: Flow diagram of the modelling process*

The process begins with the extraction of data from the database (the HYPRES database is implemented in Microsoft Access, and is detailed in Chapter 2), and is achieved with the writing of an SQL query to search for all data that possesses a particular set of desired properties (data selection criteria are detailed in Section 5.1). This is followed by the export of the dataset into Microsoft Excel. The next step is to determine which data would be relevant in the modelling (see Section 5.1), and how best to pre-process that data (see Section 6.1.2) prior to their presentation to an artificial neural network ensemble. At the modelling stage there are a number of parameters that may be altered in order to optimise the process, such as optimum

number of soil horizons (individual samples) per artificial neural network and optimum number of members in the ensemble.

Once the model has been constructed, it is then challenged with a variety of test samples. The test sample results (predictions) are then compared with expected values. The discrepancy between the predicted and expected values are analysed with a set of statistical tests and graphical representations of the model statistics, enabling the strengths/weaknesses of the modelling process and the data to be evaluated. At this stage, and indeed any other stage, conjectured improvements are made to the model and the process is started again. In this way, the development of the model may be improved at each step by learning from the many lessons on offer during the journey. At the various stages there are many factors to be considered. For instance, during pre- and post-processing we need to consider which data will yield the maximum information from the model, and how much of this data is required. The model will also require that the data be presented to it in a specific format, so it will need to be transformed in some way. During the modelling process, there are considerations such as artificial neural network structure and architecture, choice of transfer functions and training algorithms, training mode, generalisation and noise. All of these, and more, are discussed at length in this thesis.

Chapter 2 introduces the HYPRES database, and describes some of the mathematical relationships between soil survey parameters and soil hydraulic properties and, in particular, some of the more popular methods for determining and predicting them. Pedotransfer functions will be discussed here at length, including the differences between 'class' and 'continuous' pedotransfer functions. Pedotransfer function evaluation criteria are also reviewed here.

Chapter 3 describes the similarities between biological and artificial neurons, and introduces the concept and basic theory of networks of artificial neurons. The necessary background is provided to enable a sufficient understanding of artificial neural networks, their construction, training regimes and algorithms, various training issues and current testing methods, in context with the research carried out here.

Chapter 4 reviews existing work on using artificial neural networks for pedotransfer function construction, and discusses some of the drawbacks of current methods outlined in the literature. It also introduces the concept of splitting a large dataset into a number of smaller, more manageable datasets, and explains why this may be beneficial to artificial neural network models. Chapter 4 also details the two

main modes of combining artificial neural networks; ensemble and modular methods, together with the critically important concept of the bias/variance trade-off and how it may be exploited in the modelling process.

Chapter 5 demonstrates the suitability of sand, silt and clay (the soil textural components), dry bulk density and organic matter content as the artificial neural network input parameters, and water retention at field capacity as the output parameter, subject to various selection criteria. The HYPRES data is analysed to determine the statistical properties of the different parameters and texture classes.

Chapter 6 shows how the parameters of a continuous artificial neural network ensemble pedotransfer function of soil water retention at field capacity can be optimised, and the optimised model trained and tested. Sensitivity analyses are conducted on the ensemble to determine which of the input parameters are of most importance to such a model. Additionally, a pedotransfer function comprising a single artificial neural network approach is constructed, which will provide a comparison between the single and ensemble methods. Finally, two possible methods of measuring errors on individual predictions are described, both theoretically and practically, and the results of the two competing methods are discussed in conjunction with the results of statistical analyses conducted in Chapter 5.

Chapter 7 demonstrates the procedures required to optimise a class artificial neural network ensemble pedotransfer function of water retention at field capacity. Sensitivity analyses are conducted on the ensemble to determine which of the input parameters are of most importance to such a model. In addition, the two methods of establishing error bars on individual predictions detailed in Chapter 6 are employed here, and the results of these two methods are discussed in conjunction with the results of statistical analyses conducted in Chapter 5.

In Chapter 8, based on the procedures to construct optimised class and continuous artificial neural network ensemble pedotransfer functions of water retention at field capacity detailed in Chapters 6 and 7, those models are used to construct class and continuous pedotransfer functions using data from the USDA-NRCS database. Firstly, the various selection criteria detailed in Chapter 5 are applied to the available data, and the data analysed in precisely the manner discussed in Chapter 5. The class and continuous pedotransfer functions are constructed, and sensitivity analyses are conducted on the ensembles to determine which of the input parameters are of most importance. In addition, the two methods of producing error bars on individual

predictions detailed in Chapter 6 are employed here, and the results of these two competing methods are discussed in conjunction with the results of statistical analyses conducted in Section 8.1.

Chapter 9 pools together all the tests, results and analyses of those models described in Chapters 6 to 8 and direct comparisons are made here. The HYPRES class and continuous pedotransfer functions are compared, as are the USDA-NRCS class and continuous pedotransfer functions. In addition, the class pedotransfer functions for the HYPRES and USDA-NRCS databases are compared, and the same comparisons are made with the continuous pedotransfer functions. The class and continuous error models are also compared.

Chapter 10 details the research that has been conducted into constructing a pedotransfer function model of hydraulic conductivity by the ensemble model.

Chapter 11 reviews each of the steps taken in the research strategy and discusses the utility of each of these steps. Conclusions based on the research findings are presented here, and discussed in detail.

Finally, Chapter 12 indicates possible future directions that research in this topic may take.

# Chapter 2

# The HYPRES Database and Pedotransfer Functions

Measuring soil hydraulic properties, such as water content and hydraulic conductivity, by direct means can be a very time-consuming and costly exercise. Difficulties encountered have led soil scientists to develop indirect means of determining this information. This entails taking easily collected soil survey parameters, such as soil texture, bulk density and organic matter content, and forming some type of mathematical relationship with less readily available soil properties, such as water content and hydraulic conductivity. These relationships are termed 'Pedotransfer Functions'. To facilitate this process, various databases of soil hydraulic properties have been constructed. This chapter will introduce the HYPRES database, and will describe some of the mathematical relationships between soil survey parameters and soil hydraulic properties and, in particular, some of the more popular methods for determining and predicting them. The differences between 'class' and 'continuous' pedotransfer functions will be discussed, along with pedotransfer function evaluation criteria.

## 2.1 The HYPRES Database

Soil hydraulic properties, mainly saturated and unsaturated hydraulic conductivity and water retention, control the flux and storage of water at the land surface. Despite the progress that is being made in direct measurement of these properties, these techniques remain relatively time-consuming, labour intensive and expensive. Since good predictions of soil hydraulic properties instead of direct measurement may be adequate for many applications, many large databases are now being made available to support the development of predictive models of soil functioning in agricultural and environmental systems. Examples include UNSODA (Leij *et al.* 1996), HYPRES (Lilly 1997, Wösten, *et al.* 1999), WISE (Batjes 1996), and the USDA-NRCS pedon database (USDA Natural Resource Conservation Service 1994). In order to overcome a lack of data regarding hydraulic properties of European soils, Wösten *et al.* (1999) brought together existing hydraulic data that resided at 20 institutions within 12 European countries into one central database. This collaboration has resulted in HYPRES, the database of HYdraulic PRoperties of European Soils.

HYPRES Version 2.0 comprises approximately 25 Megabytes of data held in 6 separate data tables and represents 95 different soil types (as defined by the modified FAO soil legend (CEC 1985) in the 1:1,000,000 Soil Geographical Database of Europe). There are 1,791 soil profiles[1] with a total of 5,560 horizons[2]. These soil horizons are subdivided into the 6 FAO (Food and Agriculture Organisation 1990) texture classes (5 mineral[3] and 1 organic[4]) and further subdivided into 2 pedological[5] classes (topsoil[6] and subsoil[7]). Since no distinction is made between topsoil and subsoil in the organic class, this gives 11 possible soil textural/pedological classes, Table 2.1a.

| | |
|---|---|
| Topsoil | - Coarse (C)<br>- Medium (M)<br>- Medium Fine (MF)<br>- Fine (F)<br>- Very Fine (VF) |
| Subsoil | - Coarse<br>- Medium<br>- Medium Fine<br>- Fine<br>- Very Fine |
| Other | - Organic |

*Table 2.1a: Texture/pedologic classes used in HYPRES*

[1] Vertical sections of the Earth's crust showing all their different depositional layers, extending into the parent rock.

[2] According to the 19th century American geologist Joseph Le Conte, the definition of horizon reads as 'The epoch or time during which a deposit was made. The strata all over the Earth, which were formed at the same time, are said to belong to the same geological horizon.' Of course, agricultural practices render this idealogical viewpoint as a little too precise. A more useful definition of the term horizon would be 'Soil layers within the profile that are reasonably homogeneous in terms of morphological characteristics and properties (eg. colour, texture, and structure)'.

[3] A mineral soil is formed from rock or sediment, and consists mainly of sand, silt and clay, but may also contain some organic matter.

[4] Organic soil is formed from plant remains and is mainly found in swamps, peat bogs or very wet areas. Organic soil may also contain large proportions of mineral soil constituents.

[5] Pedology (from the Greek *pedon*, meaning soil) is the scientific study of soils, their origins, characteristics and uses.

[6] Topsoil – soil consisting of various mixtures of sand, silt, clay and organic matter; considered to be the nutrient-rich top layer of soil that supports plant growth. Also known as the A or the Ah horizon.

[7] Subsoil – the stratum of earth that lies immediately beneath the topsoil, but above the bedrock.

According to Wösten, *et al.* (1998), organic soils are defined as horizons with %OM[8] ≥12 when %Cl[9] = 0, with %OM ≥18 when %Cl > 60, and variable %OM when 0 < %Cl < 60 (Figure 2.1a).

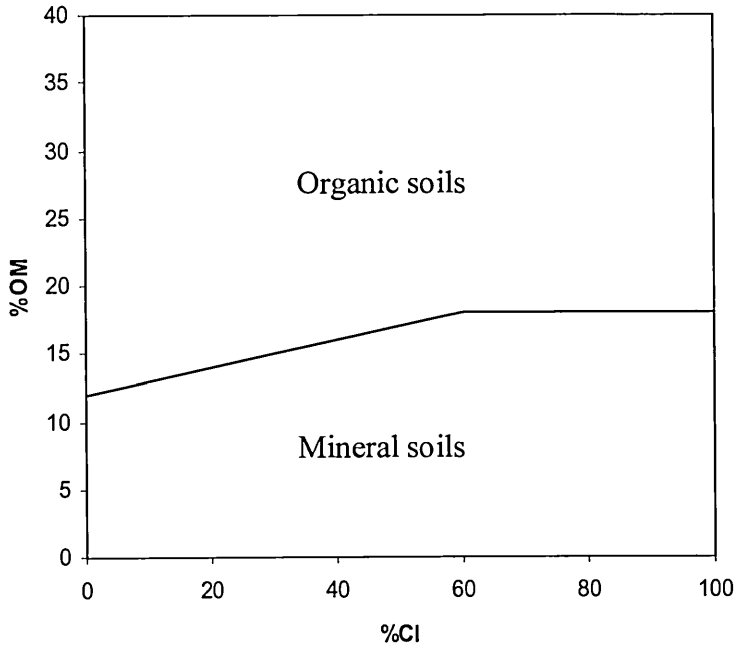



*Figure 2.1a: Distinction between organic and mineral soils in HYPRES*

Early tests showed that there are very few soil horizons in HYPRES that have %OM > 10 (less than 2% - see Table 5.1c), so in order to simplify the dividing line between organic and mineral soils (Occam's razor, Chapter 0) I decided that an organic soil is one that is defined to have 10% or more of organic matter content by volume, regardless of the %Cl content. This point will be justified in further detail in Section 5.1.

The textures of Table 2.1a are defined according to the FAO system (FAO 1990), and illustrated in the soil textural triangle of Figure 2.1b, where sand (50-2000μm), silt (2-50μm) and clay (<2μm) are defined according to the size of the soil particles.

[8] Throughout this thesis, organic matter content will be denoted as %OM.

[9] Throughout this thesis, proportions of sand, silt and clay will be denoted individually as %Sa, %Si and %Cl, respectively, and when taken together, simply as SSC.

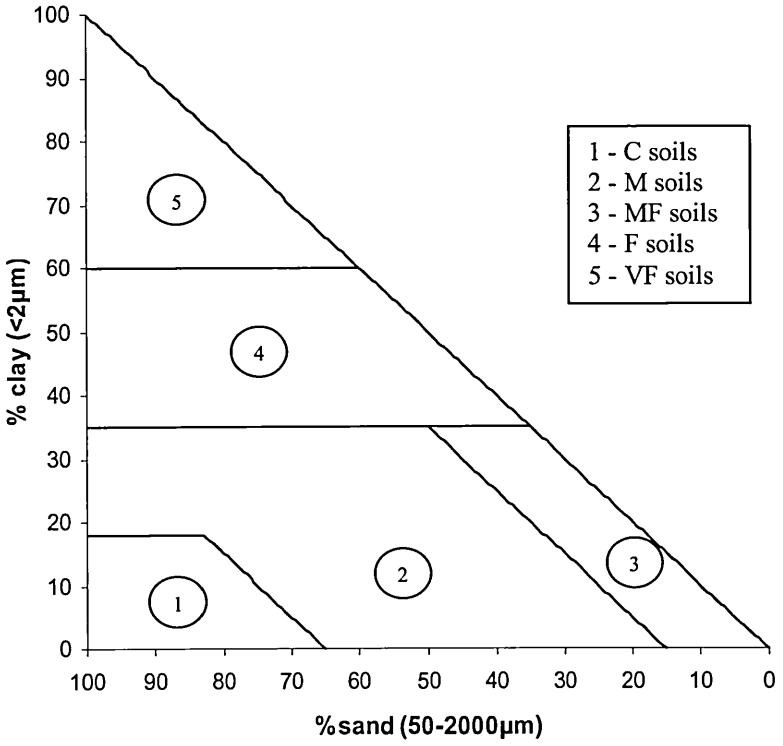



*Figure 2.1b: Illustration of the FAO (1990) definition of soil texture classes*

Soil hydraulic properties are measured by various methods. However, the measurement of pressure head values, $h$ (a parameter in Bernouilli's equation, see Section 2.2.1), is standardised to produce soil moisture retention ($\theta$), also commonly known as volumetric water content, and hydraulic conductivity ($K$) values at the following 14 pressure heads[10]: 0, -10, -20, -50, -100, -200, -250, -500, -1000, -2000, -5000, -10000, -15000 and -16000cm$H_2O$[11] for each soil horizon. Plotting $\theta$ against $h$ gives what is commonly known as the water release curve (WRC) (see Figure 2.1c for

[10] It should be noted that since HYPRES comprises data from 20 institutes in 12 countries, pressure heads are measured differently in different laboratories, conforming to different (inter)national standards. Sometimes there are differences within a country. Thus, in the construction of the database it was necessary to find some way of standardising these values. The van Genuchten equation (see equation 2.2h) was used for this purpose and, inevitably, the standardised values thus contain a degree of noise for which the fitting error is unknown. The database also contains information as to whether the value (water retention or hydraulic conductivity) at a specific pressure head is the measured value, or has been estimated.

[11] Historically, different scientific disciplines use different units of pressure to suit their particular purpose or scale of measurement. For instance, in physics, pressure is usually measured in units of $Nm^{-2}$, atm or Pa, whilst in engineering it will often be in units of mmHg or psi. In this thesis all pressures quoted will be in units of cm$H_2O$ or kPa. The conversion rate between the two is: 1kPa=10.197cm$H_2O$.

an example of this); and $K$ against $h$ gives the hydraulic conductivity characteristic (HCC).

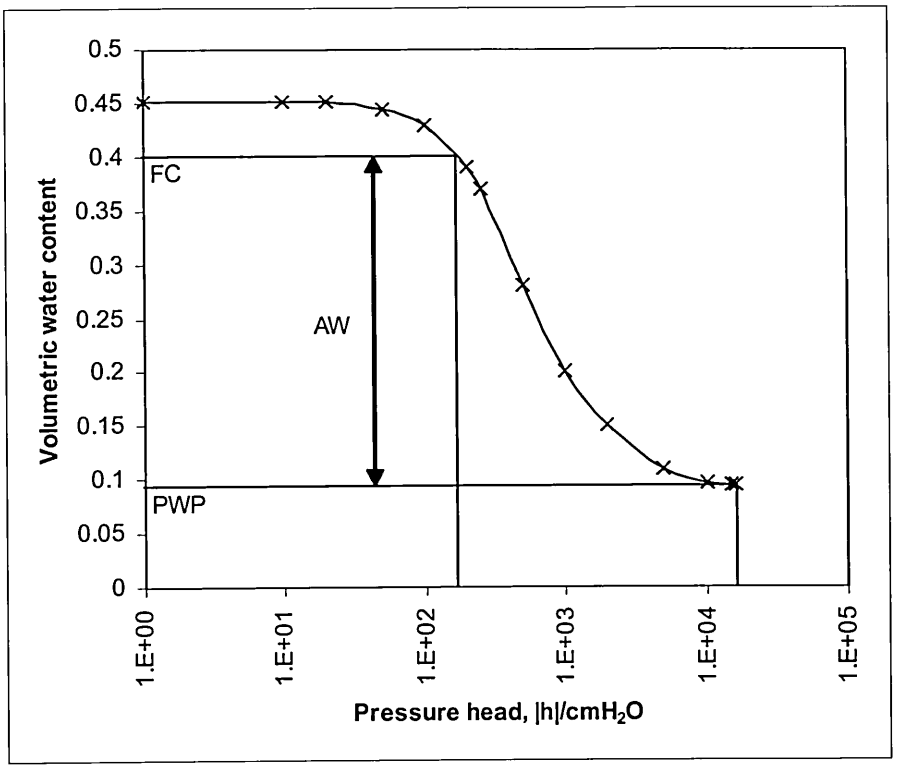



*Figure 2.1c: A typical water release curve, where FC and PWP signify Field Capacity[12] and Permanent Wilting Point[13], respectively, and AW is the Available Water[14]*

[12] The term 'field capacity' has been defined in various ways, including: (a) the maximum water holding capacity of the soil, above which all excess water drains or overflows, (b) an upper limit of water available for transpiration, (c) a water potential generally between -10 and -33kPa depending on soil texture and other properties, (d) when drainage becomes negligible after thorough wetting and (e) 2 or 3 days after a thorough wetting. Needless to say, these definitions are rather vague, and hence, the precise point of field capacity on the water release curve is similarly vague.

[13] Permanent wilting point is defined to be the point at which water molecules are held so tightly by the soil matrix that a plant cannot take up water fast enough to sustain its life. By convention this is taken to be at -1500kPa, but, as with field capacity, this value varies considerably with soil texture and other properties.

[14] Available water is defined as the water held in the root-zone between field capacity and permanent wilting point, and is the total amount of water available for the plant to withdraw from the soil.

## 2.2 Interactions Between Soil and Water

There are many forces acting within the soil-water matrix, between individual water molecules and between water molecules and soil particles. Such forces include hydrogen bonding between molecules of water, adsorption between soil particles and water molecules, and capillarity, a phenomenon that occurs at the interface between air and water. All of these interactions, and more, contribute to the emergent phenomena of water retention and hydraulic conductivity, and are discussed here in detail.

### 2.2.1 Water – Water Interactions

Water possesses extremely unusual attributes, and it is because of these that it is the most abundant molecule on the planet. Despite being a compound of two gases, water is a liquid at room temperature, where logic states that it should be a gas. Compared with other common liquids, water has unusually high melting and boiling points, viscosity and surface tension, and is also more dense as a liquid than as a solid. All of these unusual properties, and more, are due to the molecular structure and the intermolecular bonding attributes (hydrogen bonding) of water, and are discussed here in detail.

#### Molecular Structure of Water

The chemical formula for water is $H_2O$, signifying that each molecule consists of two atoms of hydrogen and one of oxygen. The hydrogen atom has a nucleus of a single proton (positively charged) which is orbited by a single electron (negatively charged). The oxygen atom has a nucleus of eight protons and eight neutrons and is orbited by eight electrons, of which six are in the outer shell. The outer shell of hydrogen is 'full' when it contains two electrons, hence it lacks one electron. Similarly, the outer shell of oxygen is full when it contains eight electrons, and thus it lacks two electrons. Hydrogen and oxygen may therefore combine in an electron-sharing arrangement, in which oxygen shares the electrons of two hydrogen atoms to complete its outer shell, whilst the two hydrogen atoms each share a single electron with the oxygen atom in order to complete theirs. This gives a central core of an oxygen atom bonded to two hydrogen atoms and two pairs of electrons. We can think of this spatially by the following analogy. Consider a two dimensional arrangement in

which a square, centred on the oxygen atom, has left-most vertices of hydrogen atoms and right-most vertices of electron pairs, bound to the oxygen across the diagonals. Now rotate the right-most vertices 90° into the third dimension so that the arrangement forms a tetrahedron. This is a spatially simplified version of the water molecule, Figure 2.2a.

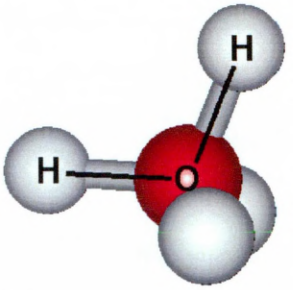



*Figure 2.2a: Water molecule showing the spatial arrangement of the hydrogen atoms and electron pairs about the central oxygen atom*

The complex interplay of forces between the positively-charged oxygen and hydrogen nuclei, the electron pairs and the hydrogen-oxygen bonds perturb the bond lengths and angles. However, the overall effect is that on one side of the oxygen (the right-most in the above example) there exists only electrons, whereas on the other side (the left-most) there exists electrons and hydrogen nuclei. This electrical assymetry gives rise to a dipole – the hydrogen nuclei side (left-most) has a small net positive charge, whilst the electron pair side (right-most) has a small net negative charge within the electrically neutral water molecule. This separation of electrical charge gives water its 'polarity'. It is this that makes water molecules mutually attractive, and is a principal reason why water 'sticks' readily on solid surfaces.

## Hydrogen Bonding

Although all hydrogen protons of water molecules are bonded to their primary oxygen atom, they are also attracted to the electron pairs of neighbouring oxygen atoms of water molecules, and may form secondary links known as hydrogen bonds, Figure 2.2b.

These hydrogen bonds are much weaker than the intra-molecular bonds (since dipole attraction is not as strong as the primary attachment between hydrogen and oxygen atoms) and act over very short distances. Hence, hydrogen bonding in liquid

water is very transitory, with tenuous associations fleetingly made. In transition from solid to liquid (melting), and from liquid to gas (evaporation), hydrogen bonds must be disrupted (whereas in freezing and condensation they are re-established). Thus, in the gaseous state, water exists mainly as monomers, free of hydrogen bonds, whereas in the liquid and solid states, their structure is increasingly governed by the phenomenon of hydrogen bonding.

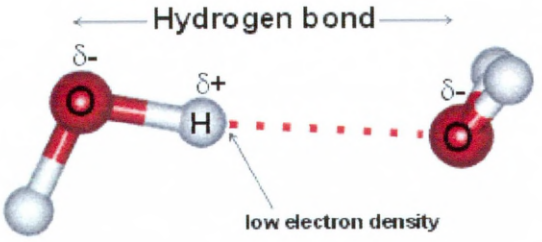



*Figure 2.2b: Hydrogen bonding between a hydrogen atom of one water molecule with the oxygen atom of another*

## Solvent Properties of Water

Due to its polar nature, water is a much better solvent than most common liquids, and is often referred to as the 'universal solvent'. Many ionic compounds, such as salt (sodium chloride – $Na^+Cl^-$), readily dissolve in water but are insoluble in non-polar liquids. Water dissolves sodium chloride because the strong electrostatic attraction between water dipoles and the $Na^+$ and $Cl^-$ ions exceeds the attraction of these ions to each other, Figure 2.2c.

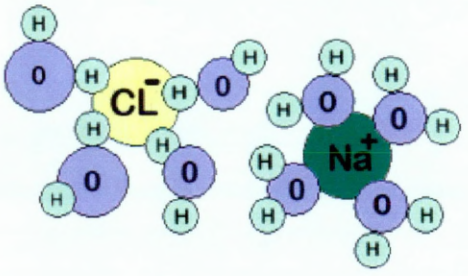



*Figure 2.2c: Hydrogen and oxygen atoms of four water molecules completing the outer shells of chlorine and sodium, respectively*

### 2.2.2 Soil – Water Interactions

Since water molecules form hydrogen bonds with other water molecules and with ionic or polar substances, it is thus reasonable to suggest that water may form hydrogen bonds with the soil matrix. In fact, they do, and the phenomenon is called 'adsorption'. Adsorption and other soil-water interactions, such as capillarity and hysteresis, are discussed in detail here.

#### Adsorption

Adsorption refers to the interactions, i.e. attraction and repulsion, between soil particles (or any other solid surface) and water. The adsorption of water on solid surfaces is mainly of an electrostatic (i.e. ionic) nature, where the polar water molecules attach to the charged faces of the solids and to the ions adsorbed on them. The interaction of the charges of the solid with the polar water molecules tends to impose an orientation on the water molecules, dictated by the charge sites on the solids. It has been conjectured (e.g. Low 1961) that the adsorbed water acquires a crystalline, ice-like structure, in which the influence of the charged site on the solid extends across several layers of water molecules. This adsorption of water is the mechanism causing the strong retention of water by clay soils, where the strength of adsorption between water and clay particles is greatest for the primary layer of water molecules. The secondary layer is attached to the primary by hydrogen bonding, and the tertiary to the secondary, etc., but the influence of the ionic site on the clay surface diminishes with distance so that beyond a few molecular layers it becomes negligibly small.

Due to its greater surface area per unit mass, as compared to sand and silt, clay is the soil fraction that has the most influence on soil behaviour. Clay particles adsorb water and hydrate, causing the soil to swell upon wetting and then to shrink upon drying. Clay particles, in contrast with the relatively inert sand and silt particles, typically carry a negative charge and when hydrated form strong ionic bonds with the surrounding water. It is mainly due to the size and adsorptive properties of sand, silt and clay that makes sandy soils loose, well-drained, well-aerated and easy to cultivate, whereas clayey soils tend to absorb and retain much more water, making them sticky when wet and hard when dry, rendering them much more difficult to cultivate.

Like clay, particles of organic matter are negatively charged. During hydration, each particle acts like an anion, capable of adsorbing various organic and inorganic constituents, including cations such as water. The cation exchange capacity[15] of organic matter is much greater per unit mass than that of clay – since it is composed mostly of carbon, hydrogen and oxygen, its charges are mainly due to the dissociation of carboxylic (-COOH) and phenolic ($C_6H_5$-OH) groups. In addition, organic matter often coagulates with clay (Theng 1973) and serves as a cementing agent, binding and stabilizing soil aggregates and thus improving soil structure.

Pores

Since the space in the soil not occupied by particles is generally continuous, it can be considered as a single pore. However, the characterisation of this pore is very difficult to define because the geometric shape of the pore space is extremely intricate and varies significantly from horizon to horizon. Nevertheless, one may consider this space as a set of discrete, interconnected pores, each of which may be classified (by the simplest classification scheme) as belonging to one of three categories of soil pore: micropores, capillary pores (see the following sub-section for an explanation of capillarity) and macropores. Figure 2.2d shows how an irregular pore may be considered as a capillary tube of equivalent cylindrical radius.

Micropores are generally less than 1μm wide, and occur typically in clayey soils.

Capillary pores are the typical pores in a medium-textured soil, ranging in width from several μm to a few mm.

Macropores may be several mm or even cm wide and are generally quite visible to the naked eye. They occur as cracks or fissures in clayey soils upon drying, as well as in all types of soils as a result of biological activity, such as burrowing animals, passage of earthworms and the presence of decayed roots. As such, macropores are usually planar or tubular in nature.

In attempting to characterise pore sizes, soil physicists often assume that soil pores are approximately cylindrical, so that an average radius may be assigned to each pore class. In reality, pores are not typically cylindrical, their widths vary along their lengths and they are variously interconnected (i.e. not distinct from one another).

[15] Cation exchange capacity (CEC) is defined as the ability of the soil to hold onto nutrients and prevent them from leaching beyond the roots, and is a measure of the adsorptive capacity of the soil. CEC is largely attributed to the clay content and the organic matter content of the soil

Thus, pore sizes based on capillary tube analogies are gross approximations, and should not be taken literally as expressing the actual radius or width of soil pores.

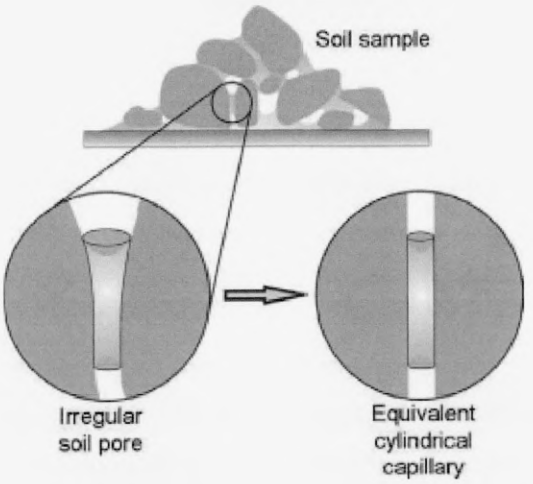



*Figure 2.2d: Illustration of the theoretical equivalency between an irregular soil pore and a capillary tube*

Surface Tension and Capillarity

A molecule inside a liquid is attracted equally in all directions by the cohesive forces of neighbouring molecules, while a surface molecule is attracted by the liquid phase to a greater extent than to the gaseous phase. This unbalanced force draws the surface molecules inward into the liquid and results in the tendency for the surface to contract, behaving as if it were covered by an elastic membrane in a constant state of tension. This is the phenomenon of surface tension, and explains why drops of liquid in air (and indeed bubbles of air in a liquid) assume the shape of a sphere – a body for which the ratio of surface area to volume is minimal, Figure 2.2e.

A capillary tube dipped in a body of water will form a meniscus as the result of adhesion at the water-glass interface and surface tension at the water-air interface, where the meniscus curvature is greater for capillary tubes of smaller bore. The pressure difference, $\Delta P$, between the water and air in the capillary tube is proportional to the surface tension, $\sigma$, and inversely proportional to the radius of curvature of the meniscus, $r$ (which is, in turn, proportional to the radius of the capillary tube), equation 2.2a.

$$\Delta P \propto \frac{\sigma}{r} \tag{2.2a}$$

The water (and its meniscus) is driven up the capillary tube by the greater pressure of the free water outside the tube at the same level (since the infinite radius of the free water is greater than that of the capillary water), Figure 2.2f. When the pressure difference between the water inside the tube and the free water reaches equilibrium with the hydrostatic pressure exerted by the column of water, the water will stop rising inside the capillary tube.

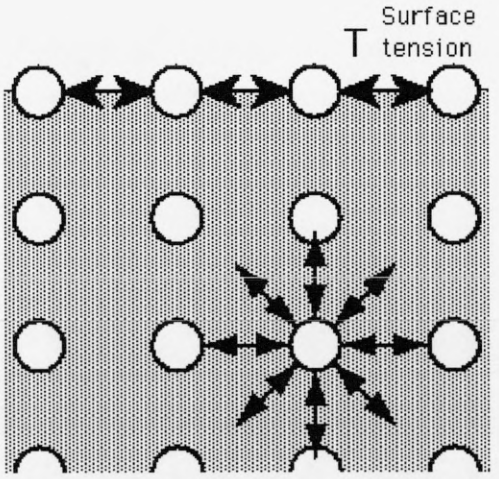



*Figure 2.2e: Illustration of the phenomenon of surface tension*

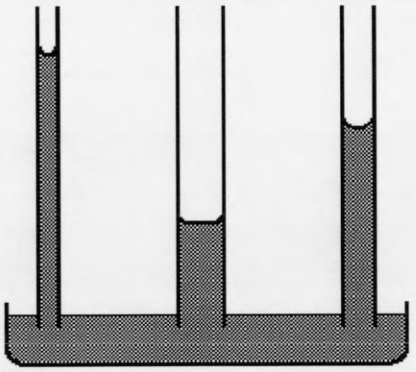


*Figure 2.2f: Illustration of the phenomenon of capillarity, demonstrating that the rise in capillary tubes is greater in smaller bores*

Now consider lifting this capillary tube above the free water. The tube will retain this water and two menisci will now result, one at each end of the vertical capillary tube. Thus, water is retained by the capillary tube, despite the effects of gravity.

It is clear to see that soil pores that have radii of similar width to capillary tubes will be subject to the laws of capillarity, and as such are described as 'capillary pores'. For capillary pores, the fluid permeating them generally obeys the laws of capillarity, and in the unsaturated state, such pores exhibit typical air-water menisci. The water retained in micropores may deviate from the simple laws of capillarity, is often discontinuous, does not participate in ordinary liquid flow phenomena, and is sometimes referred to as 'adsorbed', 'bound' or 'residual' water. When empty of water, macropores constitute barriers to capillary flow, permitting only very slow creep along their walls. When filled with water they permit very rapid flow, and although the fractional volume of macropores may be quite small, they may well have a decisive effect on drainage and the transport of solutes in saturated or near-saturated conditions.

## Hysteresis

The WRC may be obtained in two ways, by desorption, and by sorption. In desorption, one would start with a saturated soil core and, by applying increasing amounts of pressure in a step-wise manner, gradually dry the soil while taking successive measurements of released water at given values of pressure head. In sorption, one would gradually wet an initially dry soil core while reducing the pressure incrementally. Each of these methods yields a continuous curve, but the two curves will generally not coincide, Figure 2.2g.

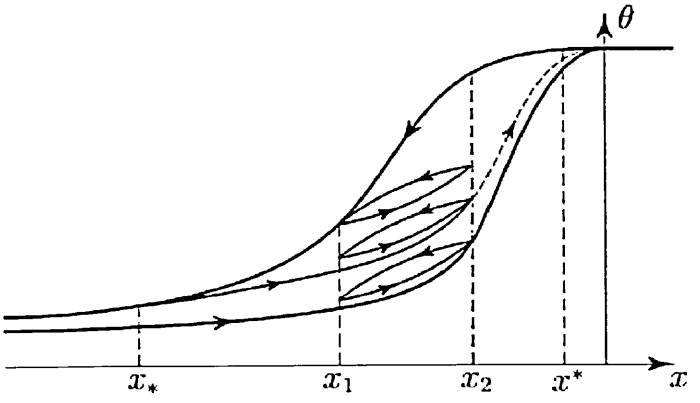



*Figure 2.2g: Illustration of the phenomenon of hysteresis*

The equilibrium soil wetness at a given pressure is greater in desorption (drying) than in sorption (wetting). This is the phenomenon of hysteresis, and is attributed to several causes:

- The encapsulation of air in 'dead-end' pores, which reduces the water content of newly wetted soil.
- The encapsulation of water in dead-end pores, which increases the water content of newly dried soil.
- The radius of curvature is greater for an advancing meniscus (sorption) than in the case of a receding one (desorption). Since the pressure difference between the pore water and pore air is inversely proportional to the radius of curvature of the meniscus, a given water content will tend to exhibit greater effects of capillarity in desorption than in sorption.
- Swelling, shrinking or ageing phenomena, which result in differential changes of soil structure, depending on the wetting and drying history of the sample.
- The geometric non-uniformity of the individual pores (which are generally irregularly shaped voids interconnected by smaller passages), resulting in the 'ink bottle' effect.

The ink bottle effect was first noted by Haines (1930), and considers a pore consisting of a relatively wide void of radius $R$, bounded by a capillary tube of width $r$, Figure 2.2h.

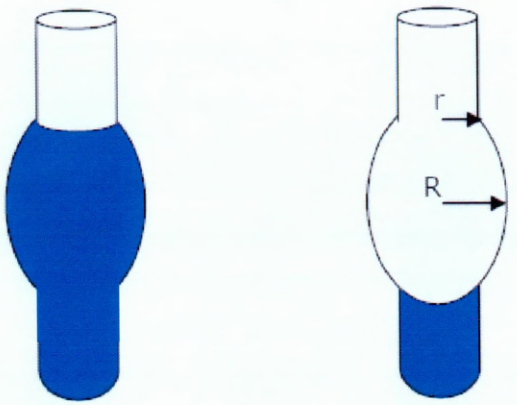



*Figure 2.2h: The ink bottle effect, illustrating the phenomenon of 'Haines jumps'*

If initially saturated, this pore will drain when the pressure difference between the air and water in the ink bottle exceeds $\Delta P_r$. For this pore to rewet, however, the pressure difference must fall to below $\Delta P_R$. Since $R > r$, it follows that $\Delta P_r > \Delta P_R$ (since $\Delta P$ is inversely proportional to the radius of curvature of the meniscus at a given point in the pore), and thus the pore fills and drains at different pressures. Clearly, desorption depends on the radii of the connecting channels, whereas sorption depends on the maximum diameter of the large pores. These discontinuous spurts of water, called 'Haines jumps', can be observed readily in coarse sands, and the hysteresis effect is in general more pronounced in coarse-textured soils for low values of pressure head, where pores may empty at an appreciably larger pressure than that at which they fill.

The two complete characteristic curves, from dryness to saturation and *vice versa*, are the main branches of the hysteretic characteristic. When a partially wetted soil begins to drain or is wetted further, the water retention follows some intermediate hysteretic curve from one main branch to the other, Figure 2.2g. The desorption curve, also known as the WRC, is the soil-moisture characteristic most commonly reported in the literature.

Flow

Most of the processes involving *in vivo* soil-water interactions occur while the soil is in an unsaturated condition. Unsaturated flow often entails changes in the state and content of soil water, involving complex relations among such variables as soil wetness, suction and conductivity, whose interrelations are further complicated by hysteresis.

The primary force acting on both saturated and unsaturated soils is that of gravity, i.e. water tends to flow downhill. However, the secondary forces acting on soil water are different for saturated than for unsaturated soils. For saturated soils, the secondary force acting is that of a pressure gradient, i.e. water tends to flow from regions of high pressure concentration to regions of low pressure concentration. For example, adding water to the top of a saturated soil core will increase the pressure differential between the top and bottom of the core, and this will have the effect of increasing the flow rate. Conversely, increasing the height of the water table below saturated soil will lead to a decreased pressure differential between the top and bottom of the soil, and the

water will consequently flow upwards until pressure equilibrium is reached. For unsaturated soils, the secondary force acting is that of matric suction, which is equivalent to a negative pressure potential. Matric suction is due to the adsorption of water and the particles of the soil matrix, capillarity and hydrogen bonding. When this suction exists, water will be drawn from zones where the hydration envelopes surrounding soil particles are thicker to where they are thinner, and from zones where the capillary menisci are less curved to where they are more strongly curved. In other words, water will flow from regions of low matric suction to regions of high matric suction.

In saturated soil, all of the pores are water-filled and conducting. The water phase is continuous and conductivity is maximal. When the soil desaturates, the larger pores (the most conductive) are the first to drain and become air-filled, thereby relegating flow to the smaller pores. The large empty pores become barriers to flow, and thus, with progressive desaturation, tortuosity increases as these larger pores are circumvented. In coarse-textured soils, water may be confined almost entirely to the contact points of the particles by capillarity, thus forming separate and discontinuous pockets of water.

Water flow in an unsaturated soil occurs as very slow creep along the walls of wide pores, and as tube flow through narrow water-filled pores. The conductive properties of unsaturated soils depend greatly on their texture and structure. At saturation, the most conductive soils are those in which large and continuous pores constitute most of the overall pore volume, whereas the least conductive are the soils in which the pore volume consists of numerous micropores. Thus, a saturated sandy soil conducts water more rapidly than a saturated clayey soil. However, the opposite is often the case in unsaturated conditions. In a soil with large pores, these pores quickly empty and become non-conductive as suction develops, thus decreasing the conductivity dramatically. In a soil with small pores, many of the pores retain and conduct water even at large values of suction, so that the hydraulic conductivity does not decrease as steeply. Since *in vivo* soil is unsaturated most of the time, hydraulic conductivity is usually greater in clayey than in sandy soils.

## 2.3 Introduction To Pedotransfer Functions

Generally, a transfer function is simply the ratio of output to input of any closed system, allowing the prediction of an output to be made when a set of inputs is presented to it. More specifically, a pedotransfer function (PTF) is a function pertaining to soil and its various parameters (Figure 2.3a).

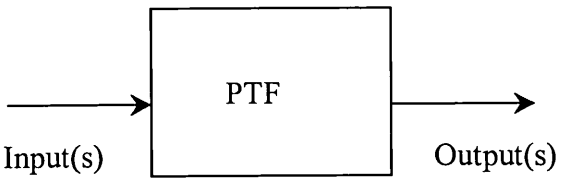



*Figure 2.3a: Illustration of a pedotransfer function*

In simple terms, when one wishes to construct a transfer function, input and output sets of data are collected from the system in question. These are then used to construct a mathematical relationship between the inputs and the outputs. This is the transfer function. The system is then ready to offer predictions of an unknown output by passing a new set of inputs to the transfer function.

A PTF (Bouma and van Lanen 1987) is a mathematical relationship between two or more routinely collected (measured) soil survey parameters, such as soil texture, bulk density and organic matter content, and soil hydraulic properties, such as moisture retention or hydraulic conductivity, that are less readily available. This relationship is used to facilitate the estimation of non-measured soil parameters, such as the WRC or HCC (or specific points along them), from one or more measured ones.

PTFs are essentially indirect methods of determining soil hydraulic properties, predicted from direct measurements recorded in soil surveys, such as percentages of sand, silt and clay, organic matter or data on particle size distribution. It is clear then that development of efficient and accurate direct measurement techniques should take place concomitant with the development of PTFs.

When considering PTFs, there are three main approaches to their development (Figure 2.3b):

- Prediction of parameters used to describe the hydraulic characteristics (Parameter or Function PTF);
- Prediction of hydraulic properties using functions derived from a simplistic soil structure model (Structure Model PTF);
- Point prediction (regression) of the hydraulic characteristic (Point PTF).

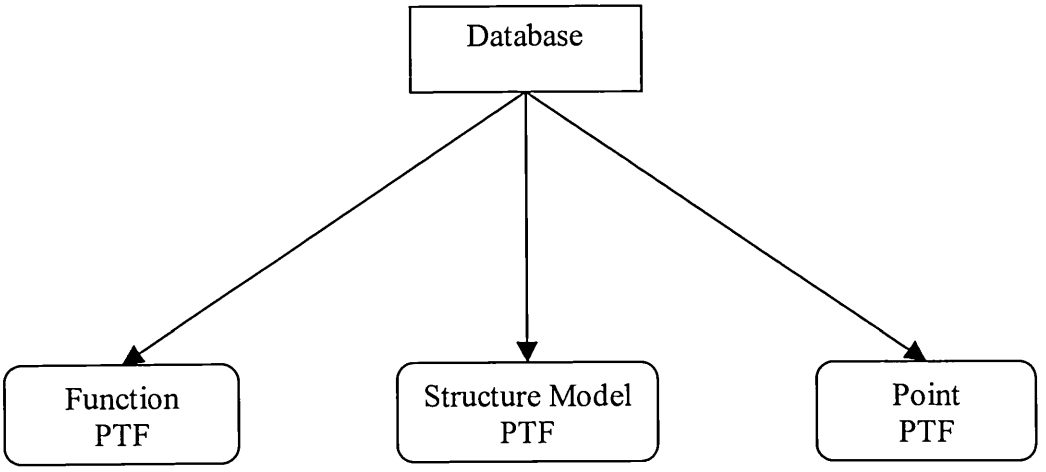



*Figure 2.3b: Schematic of the three differing approaches to PTF development*

Each of these is discussed in detail (see Sections 2.3.3 to 2.3.5), but first, it is necessary to describe the functions commonly used to describe water retention (Section 2.3.1) and hydraulic conductivity (Section 2.3.2). These are presented here in a historical context, and provide the mathematical means by which one may construct PTFs of such, whether they be function, point or structure model PTFs.

### 2.3.1 Functional Descriptors of Water Retention

The soil volumetric water content[16] is defined to be the fraction of the total volume of soil that is occupied by the water contained in the soil (e.g. see the

[16] Also known as the volume fraction of soil water, moisture content or water retention. These terms will be used interchangeably in this thesis.

Environmental Assessment Division of Argonne National Laboratory website (EAD-ANL)). Assuming that $V_l$ is the volume of the liquid phase (water) in the soil sample, $V_a$ is the amount accumulated (drained from) below the sample and that $V_t$ is the total volume of the sample, the volumetric water content, $\theta$, can then be defined as follows:

$$\theta = \frac{V_l}{V_t} = \frac{V_t - V_a}{V_t} = \frac{V_l}{V_s + V_p} \tag{2.3a}$$

where $V_s$ and $V_p$ represent the volumes of the solid phase and the pore space, respectively.

A number of functions have been used and developed to describe water retention characteristics, the most common of which are briefly discussed here.

**Bernoulli's equation**

Bernoulli's equation (Bernoulli 1738) is essentially a special case of the law of conservation of energy, and relates pressure, velocity and the height above an arbitrary reference level in the steady, energy-conserved, laminar motion of an ideal (i.e. incompressible, constant density and frictionless) fluid in a pipe, Figure 2.3c.

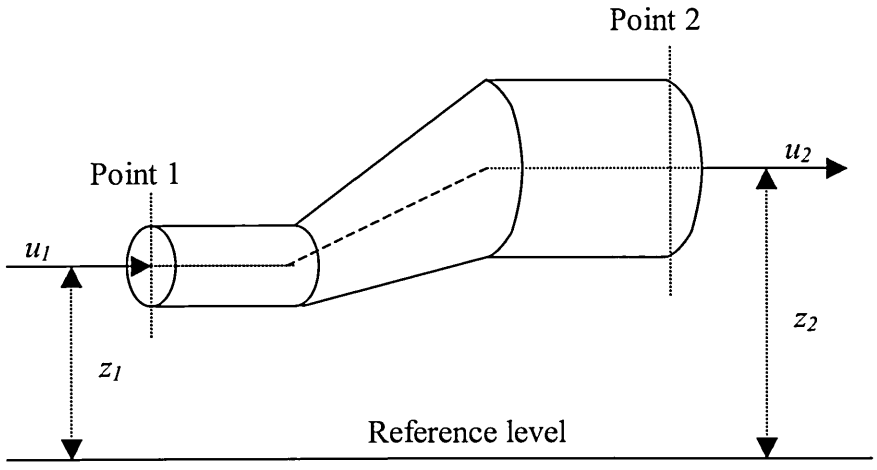



*Figure 2.3c: Illustration of fluid flow, passing points 1 and 2 in a pipe*

In its standard form, it is stated as Equation 2.3b or 2.3c.

$$\frac{u^2}{2g} + z + \frac{p}{\rho g} = C \tag{2.3b}$$

$$\frac{u_1^2}{2g} + z_1 + \frac{p_1}{\rho g} = \frac{u_2^2}{2g} + z_2 + \frac{p_2}{\rho g} \tag{2.3c}$$

where $u$ is the flow velocity, $g$ is the gravitational constant, $z$ is the height above an arbitrary reference point, $p$ is the pressure exerted by the fluid, $\rho$ is the fluid density and $C$ is a constant. In simple terms:

velocity head + elevation head + pressure head = total head

In the present context, the velocity of flow, $u$, is so small (flow velocity of water in soil is typically measured in units of hours or days) as to render the velocity head term negligible, and thus the hydraulic head $h$ is given as:

$$h = z + \frac{p}{\rho g} \tag{2.3d}$$

In hydrology, the hydraulic head $h$ is often referred to as the pressure head, and will be referred to as such throughout this thesis.

Clearly, many of the assumptions of Bernoulli's equation are violated when considering the movement of water through soil:

- Steady, laminar flow – there is no guarantee that flow of water in soils will be steady, and is certainly not laminar.
- Constant density fluid – the density of water fluctuates considerably when acted upon by changes in temperature and/or pressure.
- Frictionless flow – considerable amounts of friction will be present between water particles and the soil matrix.
- Flow in a pipe – soil is a three-dimensional matrix that resembles a number of different-bored pipes in only a simplistic manner.

- Additionally, Bernoulli's equation does not take into account the capillary action of the soil pores on the water, which is considerable.

Despite all of these shortcomings, Bernoulli's equation provides a relatively simple, if crude, means of measuring the soil volumetric water retention.

**Brooks-Corey equation**

The Brooks-Corey equation (Brooks and Corey 1964) was derived from a model in which pores are represented as a group of parallel capillary tubes with different radii, and gives a power law water retention equation:

$$S = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \left(\frac{h_b}{h}\right)^{\lambda} \tag{2.3e}$$

where $S$ is the saturation degree, and $\theta_s$ and $\theta_r$ are the saturated and residual water contents ($m^3m^{-3}$). The saturated water content pressure head $h_b$ (air entry potential), and the pore distribution index $\lambda$ are empirical fitting parameters.

Anderson *et al.* (1985) and Gregson, Hector and McGowan (1987) represented the Brooks-Corey equation in an alternative form:

$$\log h = a + b.\log\theta \tag{2.3f}$$

**van Genuchten equation**

Using a statistical pore-size model, Brutsaert (1967) published a water retention equation based on a logistic function:

$$S = \frac{1}{1 + (\alpha h)^n} \tag{2.3g}$$

where $\alpha$ is the reciprocal of the air entry potential, and $n$ is the gradient of the linear portion of the soil water content *versus* pressure head curve.

Following from Brutsaert's research, van Genuchten (1980) proposed the following, more flexible equation:

$$S = \frac{1}{\left[1 + (\alpha h)^n\right]^m} \qquad (2.3h)$$

where $m$ is most often assumed to be $m = 1 - 1/n$.

Pachepsky *et al.* (1982) and Vereecken *et al.* (1989) have developed PTFs for the parameters of the Brutsaert equation (Eq. 2.3g).

PTFs for the van Genuchten equation (Eq. 2.3h) have been developed by Wösten and van Genuchten (1988), Bachmann and Hartge (1991), Schaap and Bouten (1996), Scheinost, Sinowski and Auerswald (1997), Schaap, Leij and van Genuchten (1998), Tomasella and Hodnett (1998), Wösten *et al.* (1999) and McBratney and Bristow (1999).

### 2.3.2 Functional Descriptors of Hydraulic Conductivity

Soil water hydraulic conductivity, the proportionality constant $K$ in Darcy's law (see below), is a measure of a soil's ability to transmit water under standard atmospheric conditions and is a function of both the porous medium and the fluid flowing through it. Hydraulic conductivity is a key variable describing soil water fluxes. The entry of water into soil, the movement of water to plant roots, the flow of water to drains and wells and its evaporation are some of the examples in which hydraulic conductivity plays a decisive role.

A number of functions have been used and developed to describe the hydraulic conductivity characteristic, the most common of which are discussed briefly.

**Darcy's law**

Water movement in soil is controlled by two factors, the resistance of the soil matrix to water flow and the forces acting on each element or unit of soil water. Darcy's law (Darcy 1856), the fundamental equation describing water movement in soil, relates the flow rate to these factors, and states that water flux in porous media is the product of the saturated hydraulic conductivity and the gradient of soil water potential (hydraulic gradient). Mathematically, this is expressed as:

$$q = \frac{Q}{At} = -K_{sat} \cdot \frac{dh}{dz} \qquad (2.3i)$$

where $q$ is the flow rate (also known as the flux density), $Q$ is the quantity of water passing through cross-sectional area, $A$, perpendicular to the direction of flow in time $t$. $K_{sat}$ is the saturated hydraulic conductivity, a measure of the ease of water movement in saturated soil, $dh/dz$ is the hydraulic gradient, and $h$ and $z$ are the soil water pressure head (hydraulic head) and soil depth respectively. Saturated flow occurs when the soil water pressure is positive; that is, when the soil matric potential[17] is zero (saturated). In most soils this takes place when approximately 95% of the total pore space is filled with water.

Darcy's law is a generalized relationship for vertical laminar flow in confined homogeneous media that is porous and saturated. It shows that the volumetric flow rate is a function of the flow area, elevation, fluid pressure and the saturated hydraulic conductivity. Obvious limitations to Darcy's law include:

- Vertical laminar flow – water flows through soil in tortuous paths and is acted upon by the various forces existing in the soil matrix (i.e. capillary action, friction, etc.), and is thus not vertical or laminar. In addition, the determination of laminar flow is dependent upon the magnitude of the Reynolds number[18]. At low Reynolds numbers, viscous forces dominate, and Darcy's law is valid. At greater values of the Reynolds number, inertial forces become more important, the flow becomes nonlinearly laminar and Darcy's law cannot be accurately applied. At greater values still, flow becomes turbulent (non-linear and non-laminar) and deviations from Darcy's law become very large.
- Confined homogeneous media – *in vivo* soil is, unlike *in vitro* soil, unconfined and heterogeneous, and thus, there are clearly going to be differences between the behaviour of soil in the laboratory and in the field. The presence of such

[17] Within the soil matrix, water is affected by forces that counteract gravity. The main force involved is called the matric force, and is due to adhesion (soil particles attract water molecules to their surface) and cohesion (water molecules are also attracted to each other). The matric potential is defined as the amount of energy that is required to raise a given volume of soil water through a given vertical distance.

[18] The Reynolds number is an experimentally determined dimensionless ratio of inertial forces to viscous forces, and is used in fluid flow studies to predict the flow velocity at which turbulence will occur.

heterogeneities does not necessarily prohibit the use of Darcy's law, but rather leads to uncertainties in conclusions based on Darcy's law.

- Linear relationship between flow rate and hydraulic gradient – although this is valid in most circumstances (excluding turbulent flow), water flow through fine grained materials does not display this linearity.

**Richards' equation**

Darcy's law is concerned with water flow in saturated soil. To model flow in unsaturated soil, consideration must be given to the balance of water flowing in, to that flowing out, i.e. the principle of conservation of mass. Combining Darcy's law with the expression for conservation of mass, yields Richards' partial differential equation (Richards 1931) for water flow in unsaturated soil:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K_h\left(\frac{\partial h}{\partial z} - 1\right)\right] \tag{2.3j}$$

where $\theta$ is the volumetric water content and $K_h$ is the unsaturated hydraulic conductivity expressed as a function of soil water pressure head.

Richards' equation is essentially a mathematical model for one-dimensional vertical flow. It suffers the same drawbacks as Darcy's equation (since it is derived from it), and becomes very complex when three dimensions are considered. In addition, since it is a model of unsaturated soil, there is also the inherent problem of hysteresis to consider.

**Kozeny-Carman equation**

Saturated hydraulic conductivity ($K_{sat}$) is predicted either directly from soil properties or from measured soil water retention. One way to use information on water retention is to apply a modified version of the Kozeny-Carman equation (Carman 1937, Kozeny 1927). The Kozeny-Carman equation is represented as:

$$k = \frac{1}{16B}.\frac{d_f^2.\varepsilon^3}{(1-\varepsilon)^2} = \frac{1}{72\tau}.\frac{d_f^2.\varepsilon^3}{(1-\varepsilon)^2} \tag{2.3k}$$

where $k$ is the Kozeny-Carman constant, $B$ (related to the hydraulic radii of flow channels) is the permeability coefficient, and $d_f$ is the particle diameter. $\epsilon$ is the effective porosity and $\tau$ is the tortuosity of the flow path, and they are expressed as:

$$\varepsilon = \frac{V_{void}}{V_{total}} \text{ and } \tau = \left(\frac{L_h}{L_v}\right)^2 \quad (2.3l)$$

where $V_{void}$ and $V_{total}$ are the void space volume and the total volume, respectively, and $L_h$ and $L_v$ are the horizontal and vertical flow distances as defined in Poiseuille's law and Darcy's law, respectively (Poiseuille's and Darcy's equations govern the behaviour of horizontal and vertical fluid flow, respectively).

The modified Kozeny-Carman equation provides a relationship between porosity and permeability for a material and is designed for a packed porous bed with random, tortuous paths available for fluid flow, and is expressed as:

$$K_{sat} = B.\varepsilon^n \quad (2.3m)$$

where the exponent $n$, often set to 2, is a shape parameter of the hydraulic conductivity versus effective porosity curve.

Although the modified Kozeny-Carman equation (Eq. 2.3m) fitted data well, Ahuja *et al.* (1989), Franzmeier (1991) and Timlin, Ahuja and Williams (1996) found that the shape parameters $B$ and $n$ varied significantly from one dataset to another.

**Snyder's equation**

Substantial efforts were made to predict $K_{sat}$ by using the complete water retention curve and by considering soil pore space as a capillary bundle with various patterns of connectivity and tortuosity. Burdine (1953), Millington and Quirk (1959) and Green and Corey (1971) applied the Kozeny-Carman equation to each capillary, or to a group of capillaries. Brutsaert (1967), Alexander and Skaggs (1987) and Raats (1992) reformulated a large number of capillary bundle models as integral expressions. Snyder (1996) noted that the majority of these equations could be represented by a single expression:

$$K_{unsat} = k_0 \theta^L \left\{ \int_0^{\theta} \left( \frac{(\theta - x)^{\beta}}{[h(x)]^{\frac{2}{\delta}+\gamma}} \right) .dx \right\}^{\delta} \qquad (2.3n)$$

where $k_0$, $L$, $\beta$, $\gamma$ and $\delta$ are empirical parameters.

Equation 2.3n is mainly used as a predictor of unsaturated hydraulic conductivity ($K_{unsat}$) as a function of pressure head or water content under the condition that at least one measured value of conductivity is used. Parameters $\beta$ and $\delta$ account for the effect of pore radii variations along pores.

**Mualem's equation**

Millington and Quirk (1959) interpreted $\theta^L$ as a factor that accounts for the occurrence of continuous pores, while Mualem and Dagen (1978) interpreted it as a tortuosity factor reflecting a decrease on soil saturation. Mualem (1976) had previously represented one of the pre-cursors to Snyder's equation as:

$$K(h) = K_S \cdot \frac{\left( \left(1 + \alpha h^n\right)^{(1-1/n)} - \alpha h^{n-1} \right)^2}{\left(1 + \alpha h^n\right)^{(1-1/n)(l+2)}} \qquad (2.3o)$$

where $K(h)$ represents $K_{unsat}$ at a particular value of pressure head. This equation is used extensively by Wösten *et al.* (1999) to predict soil unsaturated hydraulic conductivity.

In addition, methods to predict $K_{unsat}$ have been proposed that are not based on the capillary bundle model. Bhatnagar, Nagarajarao and Gupta (1979) reported a linear relationship between $log(K_{unsat})$ and water content $\theta$, and suggested that $\frac{d}{d\theta}(\log K_{unsat})$ should be correlated with texture and other soil properties.

### 2.3.3 Parameter Prediction of Hydraulic Characteristics

The term Pedotransfer Function was first introduced by Bouma (1989), and defined as functions that relate different soil characteristics and properties with one another. PTFs are functional relationships that transfer available soil properties (e.g. texture, structure, organic matter content) into missing soil properties (e.g. soil hydraulic and soil chemical characteristics).

Parametric PTFs[19] are based on the assumption that the $\theta$-$h$-$K$ relationship may be described adequately by a hydraulic model (a mathematical equation such as those detailed in Sections 2.3.1 and 2.3.2 relating $h$ to $\theta$ and/or $K$), such as those formulated by Brooks and Corey (1964) and van Genuchten (1980). This approach has the advantage over point prediction methods in that it yields a continuous function of the $\theta$-$h$-$K$ relationship. However, since all of the equations of Sections 2.3.1 and 2.3.2 are essentially approximations to the functions underlying the physical processes active in the soil-water matrix, models constructed using data derived from these equations necessarily contain the same assumptions that were used in deriving the mathematical equations. Additionally, function PTFs provide an indirect relationship between measured soil survey parameters and soil hydraulic properties, with the parameters of the mathematical equation providing the intermediate link in the modelling process.

### 2.3.4 Prediction of Hydraulic Characteristics by Soil Structure Model

Bloemen (1980), Arya and Paris (1981) and Arya and Dierolf (1992) presented models that predict water retention from particle-size distribution, bulk density and particle density. Equivalent pore-size distributions are derived from these models, relating them to a distribution of water contents and associated pressure heads. Assuming water flow through cylindrical soil pores, this model is then used to predict the hydraulic conductivity.

The concept of soils consisting of a number of cylindrical pores of varying diameter is very rudimentary, and perhaps only useful in terms of considering the basic idea of water flow. More accurate predictions of hydraulic properties based on this type of model would need to provide a more precise representation of pore

[19] Parametric PTFs are also known as Function PTFs since they predict the parameters of a closed-form analytical equation (Cornelis *et al.* 2001, Acutis and Donatelli 2003). These terms will be used interchangeably in this thesis.

geometry in real soils, considering factors such as variations of radii along a pore, pore connectivity and tortuosity (Young, Crawford and Rappoldt 2001), fluid viscosity, surface tension and capillary action.

### 2.3.5 Point Prediction of Hydraulic Characteristics

Gupta and Larson (1979), Rawls, Brakensiek and Saxton (1982), Ahuja, Naney and Williams (1985) and Batjes (1996) developed PTFs in the form of regression equations predicting water retention characteristics in terms of routinely measured soil parameters, such as sand, silt, clay, etc. These functions are of the general form:

$$\theta_h = a(\%Sa) + b(\%Si) + c(\%Cl) + d(\%OM) + e(BD) + \ldots + x(X) \qquad (2.3p)$$

where %OM and BD represent the percentage by volume of organic matter content and dry bulk density[20], respectively, $\theta_h$ is the volumetric water content at pressure head value $h$, and $a$, $b$, $c$, $d$, $e$, ... and $x$ are constants. The variable $X$ refers to any other basic soil property that can be measured easily and routinely. The constants $a$ to $x$ are determined by regression of $\theta$ at various values of pressure head $h$ versus relevant soil properties.

Such PTFs are termed Point PTFs, since they estimate the water retention or hydraulic conductivity at certain values of pressure head (Cornelis *et al.* 2001, Acutis and Donatelli 2003).

This method produces fairly accurate predictions at specific points along the water retention curve, and also offers some insight as to which soil properties contribute the most towards this prediction. However, it is clear that a large number of regression equations are required to quantify the complete soil moisture retention characteristic, and this results in a rather extensive look-up table.

In addition, considering the mineral contents of soils,

$$\%Sa + \%Si + \%Cl = 100 \qquad (2.3q)$$

[20] As with sand, silt and clay, the notation of BD for percentage of dry bulk density will be consistent throughout this thesis. Also, all quoted values of BD in this thesis are in units of $g.cm^{-3}$.

Clearly, sand, silt and clay are not independent variables – strictly speaking, regression analysis should not be performed when one variable may be directly derived from the other variables.

Another problem with this method is that it is possible that the moisture retained at one pressure head value ($h$) is predicted to be less than the moisture retained at a more negative pressure head; for example, the moisture retained at –10kPa may be predicted to be less than that retained at –50kPa. This is clearly physically incorrect, but has been noted with this type of prediction.

Despite the shortcomings of point prediction PTFs, they have a significant advantage over function PTFs in that they provide a direct relationship between measured soil survey parameters and soil hydraulic properties. It is for this reason that the models detailed in this thesis (Chapters 6 and 7) are constructed using the point PTF method, directly relating measured soil survey parameters (SSC, BD and %OM) with water retention. Models relating soil survey parameters with hydraulic conductivity are discussed in Chapter 10.

## 2.4 Class Pedotransfer Functions

As a precursor to developing PTFs, a database of measured soil hydraulic characteristics can be subdivided into groups of soils, according to, for instance, taxonomy, soil moisture regime, soil temperature regime and soil textural class. This approach has the advantage that the correlation of hydraulic characteristics with other soil properties is perhaps more stable and consistent within groups of soils having similar flow processes. As a consequence, it is to be expected that more accurate PTFs can be developed for database classes as compared to the database as a whole.

The term 'class' PTFs (Wösten *et al.* 1990) refers to PTFs based on a grouping of the data, often according to the functional behaviour of different horizons but may also be according to some other criteria. Wösten *et al.* (1999) divided the soil horizons into the 11 soil textural/pedological classes of Table 2.1a, and in this thesis, the term 'class PTF' refers to soils grouped according to these 11 soil classes.

Class PTFs for the FAO texture classes were derived by Wösten *et al.* (1999) by determining the moisture contents and hydraulic conductivities at 14 arbitrary pressure heads. Since $\theta(h)$ and $K(h)$ are logarithmically distributed, the geometric

mean (GM) moisture contents and conductivities were calculated. In addition, the $\theta$ and $K$ values within one standard deviation (SD or $\sigma$) were also calculated, giving an indication of the deviation of the individual curves from the geometric mean curve. This work has been verified by replicating the work carried out by Wösten *et al.* (1999), and the water retention PTF for the coarse topsoil class is given in Figure 2.4a.

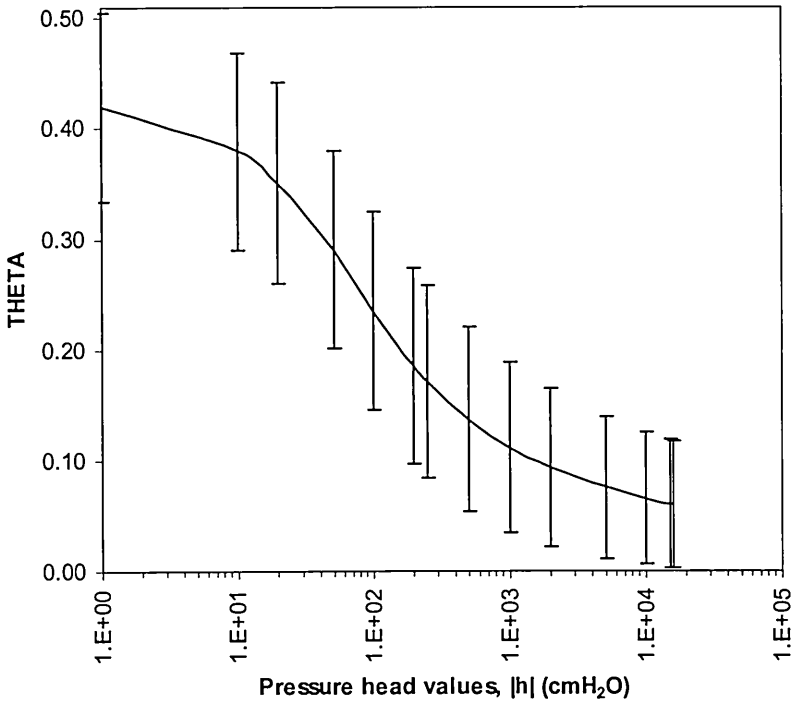



*Figure 2.4a: Coarse topsoil class PTF. Verification of geometric mean and standard deviations calculated by the method of Wösten* et al. *(1999). The curve is the GM, the upper and lower error bars are the GM + SD and GM – SD, respectively*

## 2.5 Continuous Pedotransfer Functions

The term 'continuous PTFs' (Wösten *et al.* 1990) refers to soils that have not previously been subdivided according to soil textural/pedological classes.

While class PTFs predict hydraulic characteristics for rather broadly defined soil texture class, and therefore do not provide site specific information, continuous PTFs can be applied at a regional level, for more site specific applications. Although in theory this statement may be true, Wösten *et al.* (1999) reported that the predictions of

the parameters that describe the hydraulic characteristics when using continuous PTFs are 'fairly inaccurate', having $R^2$ values ranging from 76% to as low as 12%. Equation 2.5a is an example of this, extracted from Wösten *et al.* (1999).

$$l^* = 0.0202 + 0.0006193(\%Cl)^2 - 0.001136(\%OM)^2 - 0.2316.\ln(\%OM) - 0.03544(BD)(\%Cl) + 0.00283(BD)(\%Si) + 0.00488(BD)(\%OM) \quad (2.5a)$$

($R^2 = 12\%$)

where $l$ is a parameter that determines the shape of the hydraulic conductivity curve, and $l^*$ is a transformed parameter, given by:

$$l^* = \ln\left(\frac{l+10}{10-l}\right) \quad 2.5b$$

It is clear then that alternative methods of constructing continuous PTFs should be found.

## 2.6 Pedotransfer Function Evaluation Criteria

As PTFs are predictive equations, they are routinely assessed in terms of correspondence between measured and predicted values. When the values used to test the PTF are the same as those used in its construction, the data is termed as 'seen'[21], and the accuracy of the equation is evaluated. When test values are different from the ones used to develop the PTF, the data is termed as 'unseen'[22], and the reliability of the equation is evaluated.

The same statistics are used to evaluate both accuracy and reliability. There have been many statistics used in PTF development, and the most common ones are described.

Here, $\theta_i$ denotes the actual (measured) value and $\hat{\theta}_i$ the predicted value, where the $i$th observation is counted from 1 to $N$ and $N$ is the total number of observations.

[21] Seen data is defined to be that which the model has previously encountered during the construction process.

[22] Unseen data is defined to be that which the model has not previously encountered in the construction process.

**Root Mean Squared Error**

The root mean squared error, *RMSE* (also known as the root mean squared deviation *RMSD* or the root mean squared residual *RMSR*), is used to calculate the mean prediction error, and is given as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(\theta_i - \hat{\theta}_i\right)^2}{N}} \tag{2.6a}$$

**Sum Squared Error**

The sum squared error, *SSE* (also known as the sum squared deviation *SSD* or the sum squared residual *SSR*), is used to calculate the square of the prediction error, and is given as:

$$SSE = \sum_{i=1}^{N}\left(\theta_i - \hat{\theta}_i\right)^2 \tag{2.6b}$$

**Mean Error**

The mean error, *ME*, is used to determine the size and sign of the bias of the prediction error, and is given as:

$$ME = \frac{\sum_{i=1}^{N}\left(\theta_i - \hat{\theta}_i\right)}{N} \tag{2.6c}$$

**Absolute Mean Error**

The absolute mean error, $ME_{abs}$, is used to determine the absolute size of the bias of the prediction error, and is given as:

$$ME_{abs} = \frac{\sum_{i=1}^{N}\left|\theta_i - \hat{\theta}_i\right|}{N} \tag{2.6d}$$

## Chapter Conclusions

The interactions between water and soil (the soil hydraulic properties) must be characterised in some detail in order for the flux (hydraulic conductivity) and storage (water retention) of water at the land surface to be sustainably managed and utilised. The soil hydraulic properties, therefore, play a crucial role in models dealing with the retention and transport of water and solutes. In recent years, there have been a number of databases constructed to allow the efficient modelling of the water release curve (the plot of $\theta$ *versus* $h$) and the hydraulic conductivity characteristic (the plot of $K$ *versus* $h$), and HYPRES is one such database of European soils. Modelling using such purpose-built databases allows the water retention field capacity and permanent wilting points to be predicted with relative accuracy, allowing the critically important available water to be calculated cheaply and quickly. The modelling procedure often leads to the production of what is known as a Pedotransfer Function. This is essentially a mathematical relationship between two or more relatively easily collected soil survey parameters, such as soil texture, bulk density, organic matter content and, often, less readily available soil properties such as moisture retention or hydraulic conductivity. There are three basic types of PTF construction method; function PTFs, point PTFs and prediction by soil structure model. Function PTFs are based on the assumption that a mathematical equation relating $h$ to $\theta$ and/or $K$ can be constructed. Point PTFs assume that there exists relationships between routinely measured soil parameters, such as sand, silt, clay, etc., and the soil hydraulic properties, and that these relationships may be characterised by regression equations. Soil structure model PTFs assume that soil pore spaces may be represented by a number of cylindrical capillary tubes having different diameters, thus allowing mathematical equations of cylinders to be applied to soil models. The methods with which these PTFs are constructed vary, and all have certain drawbacks.

Often, data is grouped according to some criteria, such as taxonomy, soil moisture regime, soil temperature regime and soil textural class. PTFs constructed with such groupings are known as class PTFs, and this often leads to more accurate PTFs than continuous PTFs, since the correlation of hydraulic characteristics with other soil properties is perhaps more stable and consistent within groups of soils having similar flow processes.

# Chapter 3

# Artificial Neural Networks

One possible definition of artificial neural networks is that they may be considered to be analytical systems that address problems whose solutions have not been explicitly formulated. This contrasts to classical computers and programs that are designed to solve problems whose solutions (although often complex) have been made explicit. In computing terms, artificial neural networks have a unique set of characteristics, having the ability to learn from experience, generalise from examples, and abstract essential information from noisy data. The topology of artificial neural networks is a logical structure in which nodes communicate with each other through synapses that interconnect them. This topology is imitative of the structure of biological nervous systems, and was originally motivated by a desire to try to understand the brain and to emulate some of its strengths. Latterly, artificial neural networks are being used to solve non-linear problems that often require parallel computing. Section 3.1 of this chapter describes the similarities between biological and artificial neurons, and networks built up from them. Subsequent sections then revisit Section 3.1 and describe the processes in much greater detail. Overall, Chapter 3 presents the necessary background to enable a sufficient understanding of artificial neural networks, their construction, training regimes and algorithms, various training issues and current testing methods, in context with the research carried out here.

## 3.1 Primer on Neurons and Networks

Computers can perform many operations considerably faster than a human being, and they do exactly what you tell them to do. Unfortunately, they cannot help when the problem to be solved is not fully understood, and moreover, they do not perform well when operating on noisy or incomplete data, often the only kind that is available. One answer is to use an Artificial Neural Network (ANN), a computing system that can learn on its own.

Most ANNs are computational systems, which are based to some degree on biological neurons or systems. The human brain is the ultimate example of a biological computing system, consisting of a network of approximately 100 billion interconnected neurons.

ANNs are similar in architecture to biological systems in that both use a number of (comparatively) simple, interconnected neurons. In a restricted sense artificial

neurons are simple emulations of biological neurons in that they can receive their input(s) from other artificial neurons in the ANN, perform simple operations on this data, and then pass their output(s) on to other neurons in the network.

Like biological networks, ANNs may be organised in layers such that the operations within a given layer can be executed in parallel and the operations across the layers can be executed serially.

### 3.1.1 Similarities Between Biological and Artificial Neurons

The biological neuron is composed of four basic parts: dendrites, soma, axon, and synapses. The neuron receives inputs from other sources, combines them in some way, performs an operation (usually non-linear) on the result, and passes this result on to other neurons (Figure 3.1a).

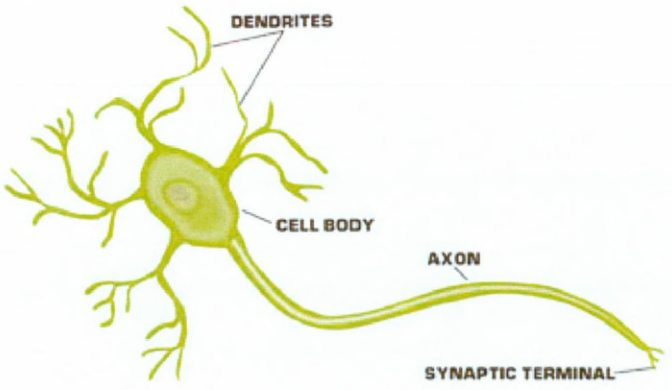



*Figure 3.1a: Illustration of a biological neuron*

The artificial neuron simulates the four basic functions of biological neurons, i.e. receives inputs, combines them, performs an operation and outputs a result. The inputs to an artificial neuron can be represented by $P_j$, where $j = 1$ to $N$ and $N$ is the total number of weights. Each of these inputs is multiplied by a connection weight, $W_j$, added to a neuronal bias term, $b$, and summed. This summation is then fed through an activation or transfer function to generate a result, which is then output to other neurons. This is illustrated in Figure 3.1b.

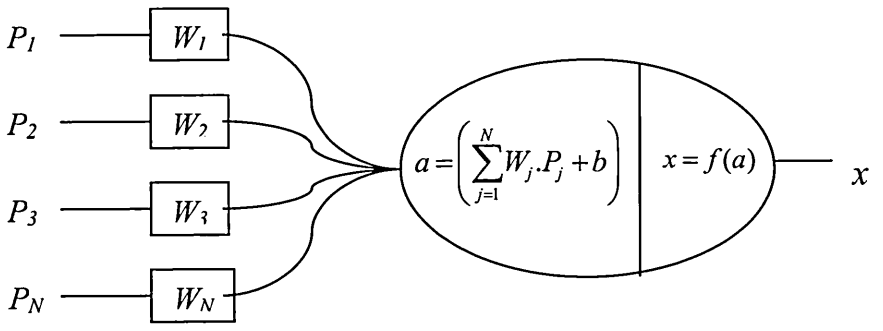



*Figure 3.1b: An illustration of an artificial neuron*

As with their biological counterparts, the real power of artificial neurons becomes evident when they are built up into networks. Although the degree of interconnectivity between biological neurons is often very complex, in simple terms, both parallel and serial processing methods can be witnessed, and it is these properties that are often exploited with networks of artificial neurons. ANNs usually have some sort of layered structure that allows parallel computations to be processed within a layer, and the results of these are passed serially on to subsequent layers (Figure 3.1c).

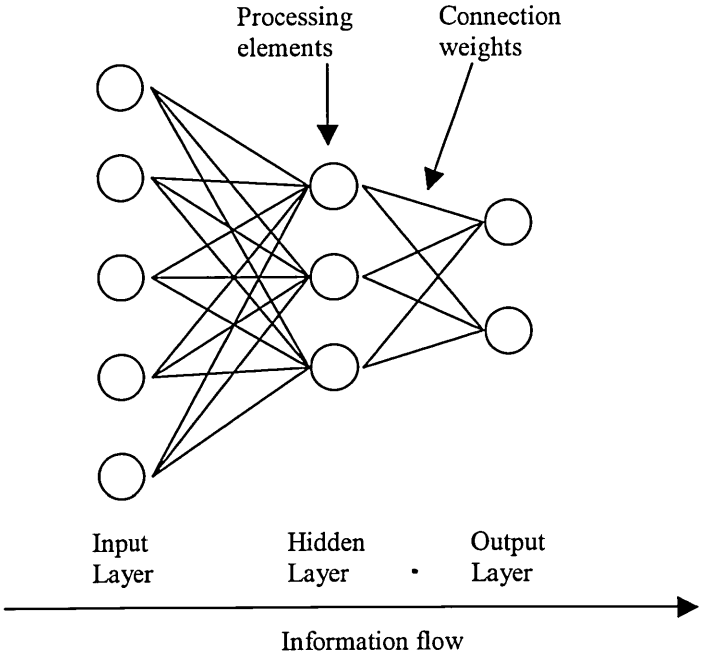



*Figure 3.1c: Schematic representation of an ANN*

## 3.2 The Biological Neuron

The human central nervous system is comprised of over $10^{12}$ interconnected neurons, the majority of these being situated in the brain, and uses biochemical reactions to receive, process and transmit information. At any time, some of these neurons are firing and the power dissipation due to this electrical activity is estimated to be in the order of 10 Watts. Monitoring the activity in the brain has shown that, even when asleep, $5x10^7$ nerve impulses per second are being relayed back and forth between the brain and other parts of the body, and this rate is increased significantly when awake (Fischer *et al.* 1987).

A biological neuron (Figure 3.1a) has a roughly spherical cell body called the soma. Electrical signals are generated in the soma and are transmitted to neighbouring neurons through an extension on the cell body called the axon. Other extensions around the cell body are the dendrites, which form a bushy tree-like structure around most (but not all) neurons, and are responsible for receiving the incoming signals generated by other neurons (Noakes *et al.* 1993).

Axons, having a length that differs from a fraction of a millimetre to a metre, joins with the cell body at a point called the axon hillock. At the other end, the axon is separated into several branches, at the very end of which the axon enlarges and forms terminal (pre-synaptic) buttons. Terminal buttons are sited in special structures called the synapses which are the junctions transmitting signals from one neuron to another. A single neuron may drive as many as $10^3$ to $10^4$ synaptic junctions.

When a nerve impulse arrives at the synapse, chemical transmitters are discharged into the synaptic cleft, the narrow gap between the terminal button of the neuron transmitting the signal and the membrane of the neuron receiving it (usually a dendrite, but may also be an axon or the cell body).

Neurons are covered with a semi-permeable membrane having a thickness of only 5nm, which is able to selectively absorb and reject ions in the intracellular fluid. The membrane acts as an ion pump to maintain a different ion concentration between the intracellular fluid and extracellular fluid. While the sodium ions ($Na^+$) are continually removed from the intracellular fluid to extracellular fluid, the potassium ions ($K^+$) are absorbed from the extracellular fluid in order to maintain an equilibrium condition. Due to the difference in the ion concentrations inside and outside, the cell membrane becomes polarised. At equilibrium the interior of the cell is -70mV (negative with

respect to the outside of the cell). This is called the Resting Potential. When not at resting potential, the potential of the interior of the cell with respect to the outside of the cell varies from –70mV to +20mV depending on the degree of polarisation.

When a neuron fires, a charge (positive or negative) is received by one of the dendrites. The strengths of all the received charges are added together and the aggregate input is then passed to the soma and on to the axon hillock. If the aggregate input is greater than the axon hillock's threshold value, then the neuron 'fires', and a constant output signal is transmitted down the axon. The output strength is unaffected by the many divisions in the axon; it reaches each terminal button with the same intensity it had at the axon hillock. This uniformity is critical in the central nervous system where small errors can magnify.

## 3.3 The Artificial Neuron

In summary, the biological neuron transmits a signal from one neuron to another through connecting nodes (synapses), and is a complex biochemical process in which specific substances (electrolytes) are released to facilitate the flow of the signal. The effect is to raise or lower the electrical potential inside the cell body of the receiving neuron, resulting in the neuron firing, if this potential reaches a threshold value.

It is this characteristic that the artificial neuron model attempts to reproduce. In 1943, neurophysiologist Warren McCulloch and mathematician Walter Pitts wrote a paper on how neurons might work (McCulloch and Pitts 1943). They modelled a simple neural network using electrical circuits, an illustration of which is shown in Figure 3.1b. This paper has since becoming the defining paper for all subsequent research on artificial neurons and their networks.

### 3.3.1 Artificial Neuron Input Function

The artificial neuron shown in Figure 3.1b has $N$ inputs (corresponding to the signals received at the dendrites in biological neurons), denoted $P_1$, $P_2$, ..., $P_j$, ..., $P_N$. Each of these neuronal inputs is assigned a weight, denoted $W_1$, $W_2$, ..., $W_j$, ..., $W_N$, respectively. Weights in the artificial model correspond to the strengths of the

synaptic connections in biological neurons. If the threshold in an artificial neuron is represented by $b$, then the activation, $a$, is given by the formula:

$$a = \left( \sum_{j=1}^{N} W_j . P_j + b \right) \tag{3.3a}$$

The inputs and the weights in artificial neurons take real values, negative for inhibition and positive for excitation. In biological neurons $b$ has a negative value, but in artificial neurons it is assigned a positive value, and is termed as the 'neuron bias', or simply as the 'bias'.

### 3.3.2 Artificial Neuron Output Function

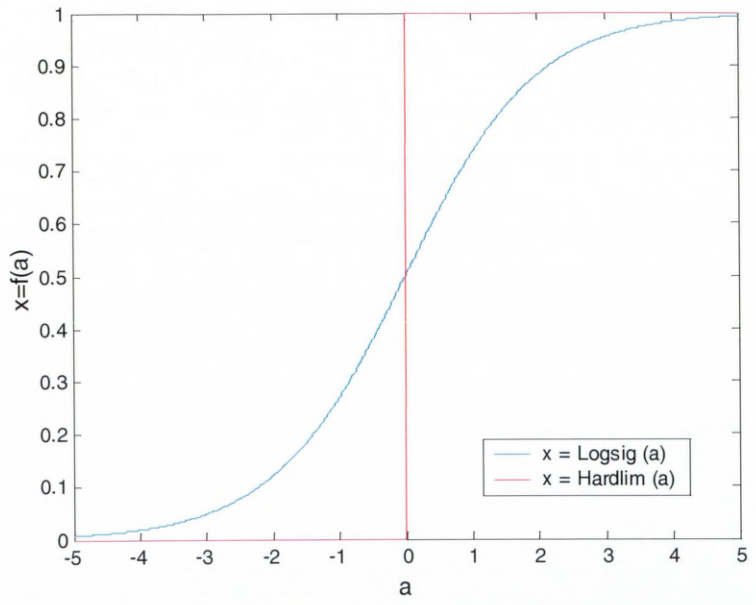
The output value of the neuron is a function of its activation, and is analogous to the firing frequency of the biological neurons:

$$x = f(a) = f\left( \sum_{j=1}^{N} W_j . P_j + b \right) \tag{3.3b}$$

The original neuron output threshold function[1] (McCulloch and Pitts 1943) may still be used in many applications. However, some of the more commonly used include linear and sigmoid functions (Figure 3.3a).

[1] The output function is commonly known as the threshold, activation or transfer function of the neuron, and these terms will be used interchangeably throughout this thesis.

| Threshold (Hardlim): | Logarithmic Sigmoid (Logsig): |
|---|---|
| $x = f(a) = \begin{cases} 0 & a \le 0 \\ 1 & a > 0 \end{cases}$ | $x = f(a) = \frac{1}{1+e^{-a}}$ |




| Linear (Purelin): | Hyperbolic Tangent Sigmoid (Tansig): |
|---|---|
| $x = f(a) = a$ | $x = f(a) = \frac{2}{1+e^{-2a}} - 1$ |

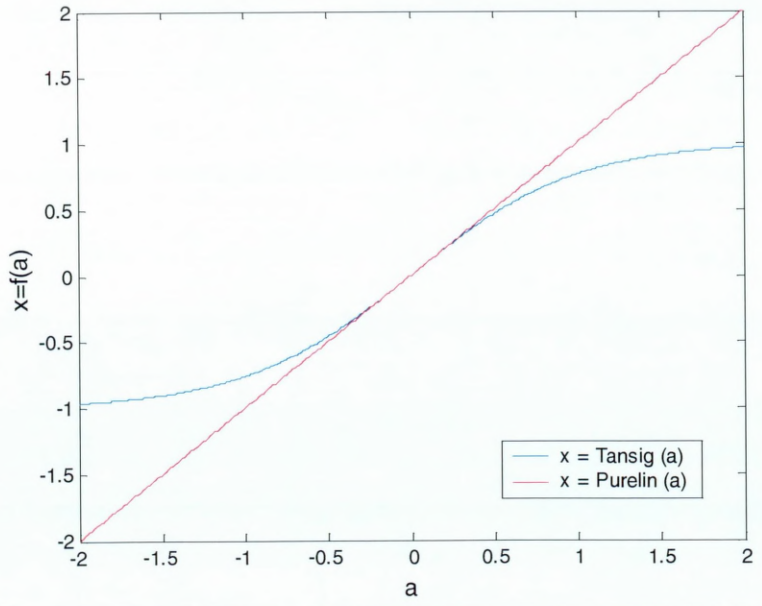



*Figure 3.3a: ANN transfer functions – hardlim and logsig (top), purelin and tansig (bottom)*

## 3.4 Networks of Artificial Neurons

While a single artificial neuron may be used to solve very simple problems, their real power becomes apparent when they are connected in networks. The simplest way of forming a network is to connect the outputs of some neurons to the inputs of others, and assign indices to the neurons to discriminate them.

If we consider $M$ layers of neurons, each layer $m$ ($m$=1 to $M$) has $N^m$ neurons (the superscript $m$ is used to denote the layer of the network referred to). The neuron $i$ of layer $m$ is connected to the $N^{m-1}$ neurons of the previous layer by a set of weights $W_{ij}^m$ ($j$=1 to $N^{m-1}$) and biases $b_i^m$.

The output $x_i^m$ of a neuron $i$ belonging to a layer $m$, is a function $f_i^m$ of the weighted sum of the outputs coming from the neurons in the previous layer $a_i^m$:

$$x_i^m = f_i^m\left(a_i^m\right) = f_i^m\left(\sum_{j=1}^{N^{m-1}}\left(W_{ij}^m.P_j^{m-1} + b_i^m\right)\right) \tag{3.4a}$$

where $P_j^{m-1}$ constitutes the outputs of the $N^{m-1}$ neurons of the previous layer.

This equation has become standard in the field of artificial neural networks and describes how data is passed from neuron to neuron within a network.

### 3.4.1 Network Architectures

Biological neural networks are much more complicated than the mathematical models we use for ANNs, and while there is no universally accepted definition of an ANN, most researchers in the field would agree that an ANN is a network of many simple processors or units, each having a small amount of local memory. The units are connected by communication channels or connections that carry encoded numeric data, and operate only on their local data and on the inputs they receive via the connections.

Some ANNs are models of biological neural networks, but historically, much of the inspiration for the field of ANNs came from the desire to produce artificial systems capable of 'intelligent' computations similar to those that the human brain routinely performs, and thereby possibly to enhance our understanding of the human brain (Ripley 1994).

Most ANNs do not program or compute in the way that conventional computers do (i.e. explicit computation), but 'learn' by examples that are passed to them by a 'teacher', a concept known as training (Mitchell 1997). The network learns by adjusting the weights of connections on the basis of data presented to it, and when trained exhibits some capability for generalisation beyond the training data.

Virtually all ANN architectures conform to one of three models (Ripley 1994) – an ANN having a single layer of neurons, multiple layers of layers, or forming a recurrent network. All of these architectures will now be briefly discussed.

## Single Layer of Neurons

A single layer ANN consists of a data input layer and a single (output) layer of neurons, where each data input layer is connected to each neuron by a single layer of weights. Each data input is connected to each neuron in the output layer by the associated weights. An example of this is shown in Figure 3.4a.

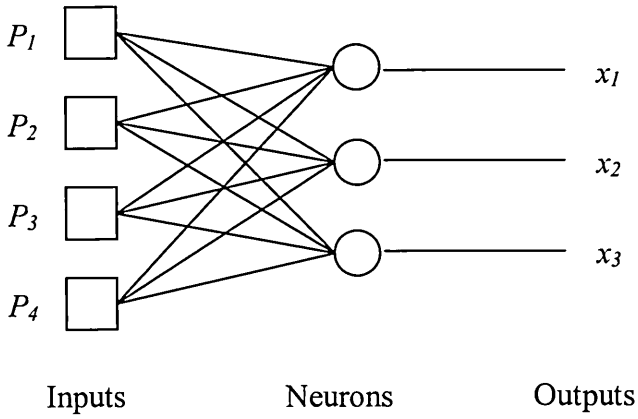



*Figure 3.4a: A single layer of three neurons, with four input parameters*

None of the neurons are connected to other neurons. It is not uncommon for the number of inputs to a layer to be different from the number of neurons, while the number of output neurons is usually specific to the given problem. In the example of Figure 3.4a, the solution is required to fit into one of three categories, $x_1$, $x_2$ or $x_3$, and so the number of output neurons is chosen to be three.

The first significant artificial neural systems with modifiable elements were termed 'perceptrons' (Rosenblatt 1958).

Multiple Layers of Neurons

A multiple layer ANN contains one or more layers of hidden neurons, sited between the input and output neuron layers, that are not directly accessible from both the input and output sides of the network (Figure 3.1c). Each layer of neurons is connected to the previous layer by the associated weights in the same way as the single layer network type, and the signal is fed forward through the layers in a forward direction only. An example of such an ANN is shown in Figure 3.4b. For this reason, an ANN having multiple layers of neurons is often termed a 'feed forward' network (FF-ANN). An FF-ANN may also be termed a 'multi-layer perceptron' (MLP-ANN).

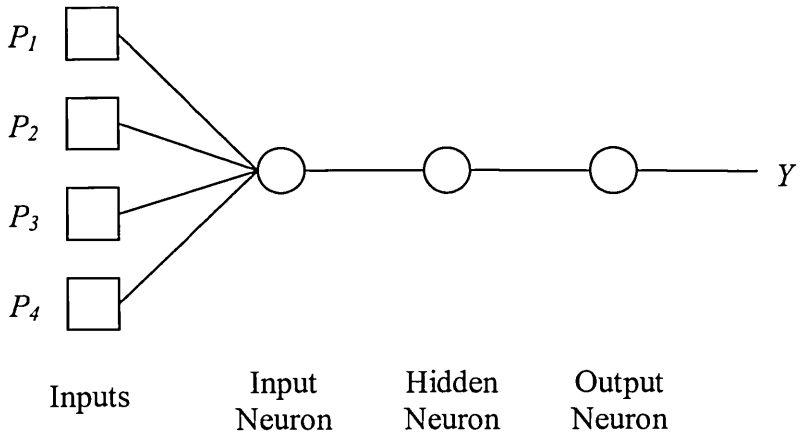



*Figure 3.4b: A three-layered MLP with a single neuron in each layer*

ANNs utilising multiple layers of neurons can solve more complicated problems than can single layered nets due to their increased complexity (Minsky 1969).

FF-ANNs are suited to a wide range of problems, including classification (Patuwo *et al.* 1993), prediction (Veelenturf 1995), pattern recognition (Patuwo *et al.* 1993), interpolation (DTI 1994) and function approximation/mapping problems (DTI 1994) which are tolerant of some imprecision, but to which hard and fast rules cannot easily be applied.

As an example of how information is passed from the input to the output stages of an MLP, consider the three-layered ANN of Figure 3.4b where there is a single neuron in each layer. All transfer functions are purelin, all weights are equal to 1 and

all biases are equal to –1. Allowing the inputs of the MLP to be $P_1 = 1$, $P_2 = 2$, $P_3 = 3$ and $P_4 = 4$, what would be the output, $x_1$, of the input neuron?

The input of the input neuron is: $a_1 = (WP_1 + b) + (WP_2 + b) + (WP_3 + b) + (WP_4 + b)$

Therefore: $a_1 = ((1 \times 1) - 1) + ((1 \times 2) - 1) + ((1 \times 3) - 1) + ((1 \times 4) - 1) = 6$

Since the output of a purelin transfer function is equal to its input, then the output of the input neuron is:

$x_1 = 6$

This is then passed to the hidden neuron, so the input to, and therefore the output from, the hidden neuron is given by: $x_H = ((W \times 6) + b) = ((1 \times 6) - 1) = 5$

And again, this is passed to the output neuron, so the output at $Y$ is:

$Y = ((W \times 5) + b) = ((1 \times 5) - 1) = 4$

## Recurrent Networks

Recurrent ANNs distinguish themselves from FF-ANNs in that they contain one or more feedback loops (some of the outputs are connected to the inputs) that can be of a local or global kind, see Figure 3.4c for an example of this.

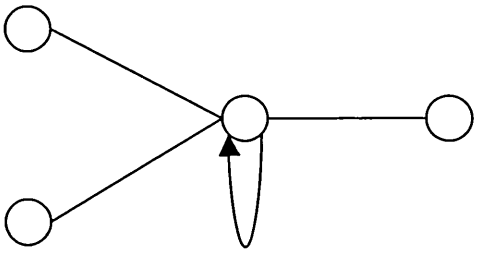



*Figure 3.4c: A recurrent ANN, with feedback loop on the hidden neuron*

In the example of Figure 3.4c, at a time $t$, the input to the hidden neuron is a weighted sum of the outputs of the input neurons at time $t$ and the output of the hidden neuron at time $t$-1. The output of the hidden neuron is thus a function of its past behaviour, and as such, recurrent ANNs are able to store information about time, and are particularly suitable for forecasting time-series applications (Ripley 1994, DTI 1994). It can be seen that once we allow feedback connections, the ANN topology becomes very free and we can connect any unit to any other, even to itself.

As an example, let us consider the state of the hidden neuron. Assume that the hidden neuron is of the purelin type, and let all weights be equal to 1, the bias equal to –1 and all inputs passing from the input neurons to the hidden neuron be equal to 0.5. What would be the output from the hidden neuron (the signal passed to the output neuron) at time points $t = 0$, 1 and 2?

Firstly, in order to calculate the input to the hidden neuron at $t = 0$, it is necessary to calculate its output at $t = 0$. Since at $t = 0$ there is no signal passing through the hidden neuron, its output is zero.

The input to the hidden neuron is thus: $a_{t=0} = (W.P+b)+(W.P+b)+(W\times 0+b)$

Therefore: $a_{t=0} = ((1\times 0.5)-1)+((1\times 0.5)-1)+((1\times 0)-1) = -2$

Since the transfer function is purelin, the output is equal to the input, so the output of the hidden neuron at time $t = 0$ is:

$x_{t=0} = -2$

At time $t = 1$, this output is then fed back into the hidden neuron to provide an input.

The input (and therefore the output) to the hidden neuron at time $t = 1$ is given by:

$x_{t=1} = a_{t=1} = (W.P+b)+(W.P+b)+((W\times -2)+b)$

Therefore: $x_{t=1} = a_{t=1} = ((1\times 0.5)-1)+((1\times 0.5)-1)+((1\times -2)-1) = -4$

Similarly, at time $t = 2$, this output is then fed back into the hidden neuron to provide an input, and hence the output of the hidden neuron at time $t = 2$ is:

$x_{t=2} = a_{t=2} = (W.P+b)+(W.P+b)+((W\times -4)+b)$

Therefore: $x_{t=2} = a_{t=2} = \big((1\times 0.5)-1\big)+\big((1\times 0.5)-1\big)+\big((1\times -4)-1\big) = -6$

This example has been carefully structured so that the state of the hidden neuron would not change over time *if the feedback is not included* (all weights, biases and inputs remain constant). When it is included, as time progresses, the feedback element becomes more negative and, as a result, the output of the hidden neuron becomes more negative. Hence it is being influenced by its past behaviour.

## 3.5 Training an Artificial Neural Network

Although virtually all ANNs fall into one of the three architectural classes, there are no restrictions on the structure of any given ANN in terms of the number of hidden layers or the number of neurons in each layer. It is clear then, that there are many different kinds of ANNs, each one constructed to fit the demands of a given problem. However, a further classification that can be applied to ANNs is the type of training regime that has been applied to them. The training or learning process consists of presenting the ANN with example data and then adjusting the ANN's internal weights until the desired ANN response is obtained. The method used to adjust the weights is known as the 'training algorithm'. There are a number of training algorithms, some of which only apply to certain types of ANN, and some of the more popular algorithms are discussed in Section 3.5.2. There are two basic classes of training algorithms; 'supervised' and 'unsupervised', and these are briefly discussed in Section 3.5.1.

### 3.5.1 Training Regimes

#### Supervised Training

In Supervised Training, training is accomplished by presenting the ANN with a set of training patterns, each having an associated target output pattern. The weights are then adjusted iteratively according to a learning algorithm in order to minimise the difference between the actual ANN output and the expected ANN output (Mitchell

1997). In conceptual terms, there is a 'teacher' who in the learning phase 'tells' the net how well it performs or what the correct behaviour would have been.

Supervised ANNs are particularly suited to prediction-based problems (Veelenturf 1995), and as such will be used in modelling PTFs.

**How a Supervised ANN works**

The data may be represented as a number of 'input patterns', designated $IP_n$, which are to be matched to an equal number of 'output patterns', $OP_n$, where $n$ represents a given sample ($n$ = 1 to $N$, where $N$ is the total sample size). In our case, an input pattern may be made up of a number of parameters such as soil texture (SSC), BD and %OM, and an output pattern could consist of one or more parameters, such as the soil water retention or hydraulic conductivity at a given value of pressure head (Table 3.5a).

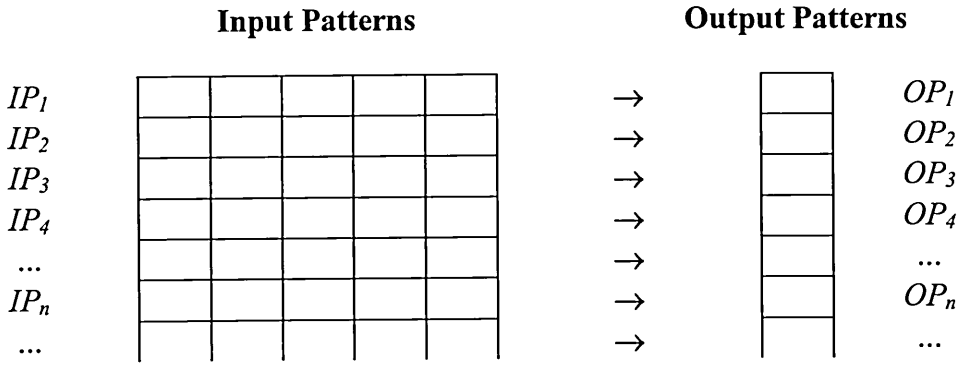



*Table 3.5a: Example of a set of input:output patterns for a supervised ANN*

A supervised ANN then uses these input:output patterns in the following way:

0. Based on the architecture, the ANN chooses random initial weights and biases of all neurons prior to the introduction of any data.
1. Input pattern $IP_1$ is passed to all of the input neurons, and the ANN updates its estimate of the weights and biases. In this way, the signal is fed forward through the network of neurons to the output stage.
2. The discrepancy between the expected and the predicted output pattern gives an error that is fed backwards through the ANN. This allows the ANN to 'learn'

from this discrepancy, enabling subsequent weight and bias estimates to be improved.

3. Steps 1 to 2 are repeated for all input:output patterns.
4. Steps 1 to 3 constitute one iteration (epoch), and are repeated until the ANN can match all input patterns with all the output patterns to within a specified error goal (that is, the average error over all input patterns).

Unsupervised Training

In Unsupervised Training, the ANN is autonomous and just inspects the data it is presented with, determines some of the properties of the dataset, and learns to reflect these properties in its output (Warner and Misra 1996). Typically, the ANN learns to cluster data into groups that exhibit similar statistical properties; what exactly these properties are depends on the particular network model and training algorithm. In contrast to Supervised learning schemes, the training set for Unsupervised learning comprises only inputs (training cases without target outputs). The training algorithm for this type of ANN is used to adjust the ANN's weights so that similar inputs consistently result in the same output.

This type of learning is often referred to as 'self-organising'.

Reinforcement Training

Reinforcement Training is a combination of supervised and unsupervised training, in that restricted information is provided about the data during training. This information consists merely of a statement as to whether the response associated with a given data set is 'good' or 'bad'. This makes the learning algorithm more effective than purely unsupervised training, but without the *a priori* assumptions that the supervised regime imbibes from the investigator (Hrycej 1992).

### 3.5.2 Training Algorithms

There are literally hundreds of different types of training algorithm, a comprehensive review of which will not be given here, instead, a selection of appropriate algorithms are introduced and discussed. Additionally, the order in which

the algorithms appear are a reflection of the evolution of my learning process with respect to training algorithms of supervised ANNs.

## Delta Rule

Bernard Widrow and Marcian Hoff (Widrow and Hoff 1960) developed a learning rule for a single layer ANN that is now known as the Delta Rule, least mean squares (LMS) rule, or the Widrow-Hoff law.

Widrow and Hoff (1960) showed mathematically that the configuration of weights that the system generates will tend to minimise the sum of the squares of the errors between the test inputs and the responses, with the efficiency of this minimisation being a function of how closely the test and training sets match. The Delta Rule is most often cast as:

$$\Delta w = \eta[d - y]x \tag{3.5a}$$

where $d$ and $y$ are the target and expected outputs, respectively, and $x$ and $\Delta w$ are the inputs and change in synaptic weights, respectively, and the symbol $\eta$ represents a positive learning rate constant.

This Law is an example of a 'gradient descent' algorithm, where each correction that is made improves the network accuracy through descent along the gradient of an error surface.

## Backpropagation

The Backpropagation (BP) Rule evolved from the Delta Rule due to the latter's inability to solve simple problems and the lack of a general method of training a multilayer network (Fausett 1994), and caused a major breakthrough in neural network research. Werbos (1974) further hypothesized a chemical flow backwards from neuron to neuron, opposite to the forward direction of electrical excitation, to establish the causal pathways. This prompted Werbos (1974) to investigate the backpropagation of errors through a feed-forward electronic network as a model of the learning process. In principle the rule is very simple: calculate the error made by

the network and propagate it back through the network layers. This back-propagated error is used to update the weights.

A BP network typically starts out with random weightings. It is then exposed to a training set of input data accompanied by the correct output corresponding to each input case. As the training proceeds, the network's weights are incrementally adjusted by the gradient descent approach until it is responding accurately (in this sense, the accuracy of the ANN is determined according to the investigator's requirements, and is an adjustable parameter known as the 'error goal'). The ANN is then ready to categorise inputs whose outputs are unknown, according to the training it has received.

The error surface that is traversed in the gradient descent approach (i.e. the difference between the current weights and the optimum weights required for accurate training, defined by some mathematical algorithm) reflects an enormous amount of information involving all possible synaptic weights and all possible input vectors. It is generally very complex and irregular, containing peaks, troughs, ridges, valleys and slopes. The goal of the BP paradigm is to locate, via gradient descent, a trough or minimum (ideally a global minimum) in the error surface. In other words, the BP paradigm is an explorer looking for a trough or valley in a fixed but very complex multidimensional landscape (Chester 1993). This trough is the minimum error, and its location is the set of synaptic weights that, when averaged over the population of input patterns, leads to this minimum.

### Backpropagation with Momentum

The existence of local minima on the error surface is compounded by the existence of multiple global minima, since many different weight permutations may lead to the same input-output transformation.

In theory, weight changes that occur in the course of training should be infinitesimally small in order to represent true gradient descent (Rumelhart *et al.* 1986). In reality, they need to be altered in finite steps, and the constant of proportionality, $\eta$ (see Equation 3.5a), also known as the learning rate, sets a scale for the size of these steps. However, problems arise as the learning rate is applied in different regions of the error surface. When the surface is relatively flat, the gradient descent proceeds very slowly, and when the surface is steep, the changes occur

rapidly and may result in oscillation about a minimum. Clearly, attempts to optimise the learning rate for any given gradient are delicate and fraught with problems.

Rumelhart *et al.* (1986) proposed the introduction of a momentum term in the BP algorithm – an inertial leftover from cycle ($t$-1) of weight changes just prior to the current cycle $t$. This is designed to smooth the transitions, the weight adjustments being a combination of the current weight adjustment and the weight change from the previous step. A useful consequence of BP with momentum is that training will often have the necessary momentum to pass through local minima, that BP without momentum would ordinarily get trapped in, and onto a global minimum that lies beyond.

### Levenberg-Marquardt

The Levenberg-Marquardt (LM) algorithm (Levenberg 1944; Marquardt 1963) has been designed specifically for minimizing a sum-of-squares error, and is widely acknowledged to provide the fastest convergence on function approximation problems for ANNs that contain up to a few hundred weights (Demuth and Beale, 2001).

The advantage over BP is especially noticeable if very accurate training is required, and in many cases, LM is able to obtain lower mean square errors than any other algorithm.

However, just because it it is quick and accurate, does this necessarily mean that it is the best training algorithm to use? To answer this question, we need to examine what we mean by the word 'best'.

One critical issue in developing an ANN is generalisation: how well will the ANN make predictions for unseen data? If an ANN is required to model three test data points, A, B and C as shown in Figure 3.5a, which of the lines of fit produce the smallest error, the straight line or the sinusoid?

Since both the straight line and the sinusoid pass through the three data points, they both produce the same size error (i.e. zero) and both perform equally well when tested with seen data. It is clear, however, that they will perform very differently when tested with unseen data.

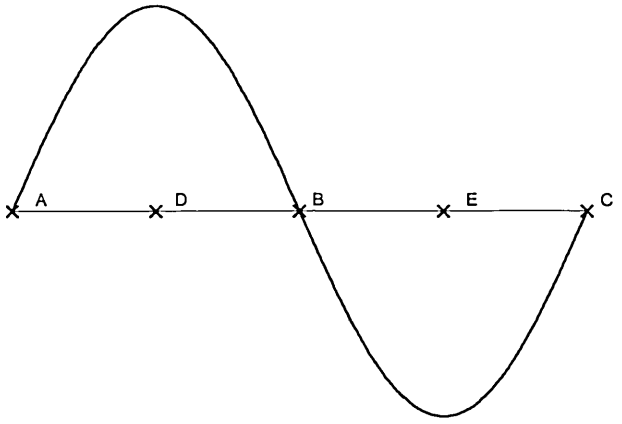



*Figure 3.5a: Illustration of the need for a validation set when using the LM algorithm*

Ideally, an ANN should work as well on unseen data as it did on the training data. In training, a normal part of the procedure is to test the ANNs performance on what is known as a 'validation set[2]', i.e. a set of instances of the input data that is different from the training set. Typically, as training proceeds the errors on the training set decrease, indicating that the ANN is learning from the training dataset. However, the errors on the validation set will decrease until the ANN is no longer learning from the training dataset, but is learning the noise inherent in the training data, at this point, although errors on seen data are continuing to decrease, errors on unseen data begin to increase (Figure 3.5b). It is at this point that training should be stopped (a process known as 'early stopping', see Section 3.6.5), since the ANN is now trained optimally.

[2] In the ANN community, there is much mis-use of the word 'validation'. I am very much of the mind that portions of a dataset should be referred to as 'training', 'validation' and 'testing'. The training set is the portion of data that is presented for the ANN to learn from, whilst the validation set is used to evaluate the performance of the ANN *at the training stage*, but is distinct from, and is not used as part of, the training set. The test set should be presented to the optimised ANN to evaluate its final performance. The terms training, validation and testing will be used in this context throughout this thesis.

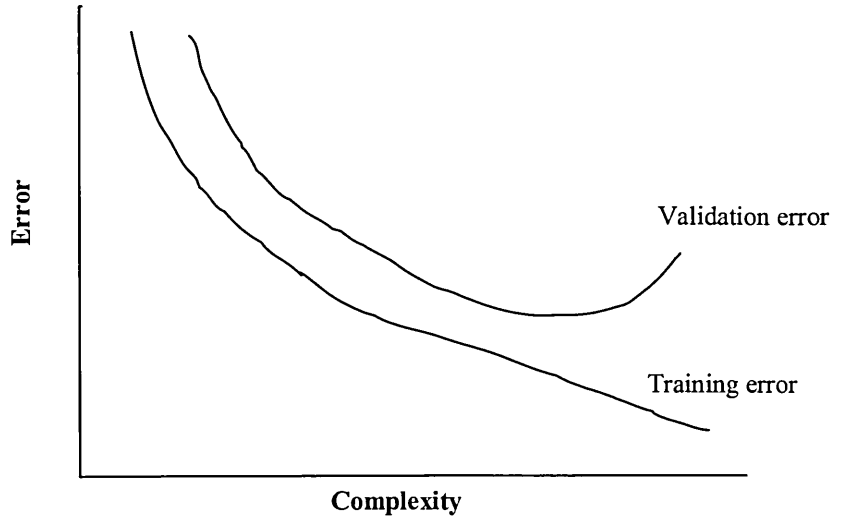



*Figure 3.5b: Relationship of complexity to test results for a validation set*

Now consider passing a further two data points D and E to the ANN during the training process as a validation set (Figure 3.5a), then the ANN will have a much better ability to approximate the underlying function of the data. Of course the disadvantage of this is that, when training, not only do we have to set aside a portion of the available data for testing, but we also have to leave out a set for validation. A further disadvantage is that the ANN is still trying to learn the data, and not the function. If the data points A to E are all subject to errors and noise, the LM algorithm will not take account of this, but will simply learn the data points regardless of what the underlying function may be.

The critical point here is that the *data* themselves are not important, only the *function* underlying the data.

## Bayesian Regularisation

The two principal approaches to the problem of inference (predictions made with unseen data) are known as *maximum likelihood* and *Bayesian inference*. Although the methods often lead to similar results their conceptual basis is significantly different.

Maximum likelihood seeks to find the optimum value for the parameters by maximising a likelihood function derived from the training data, the optimum value being the single most likely value of a parameter to give rise to the observed data. By contrast, in the Bayesian approach, the parameters are described by a probability

distribution[3], and the optimum value now becomes the central point in a distribution of probable values. Unlike the maximum likelihood method, the Bayesian approach does not involve setting the parameters to specific values, but rather allows us to express our uncertainty in the values of the parameters by a probability density function. Before we observe the data, the parameters are described by a 'prior' probability density, which is typically very broad to reflect the fact we have little idea of what values the parameter should take, Figure 3.5c.

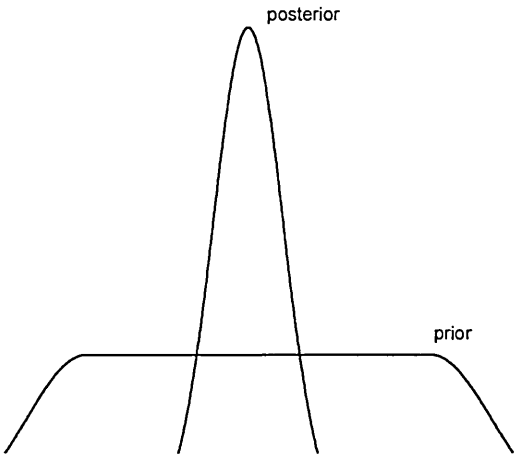



*Figure 3.5c: Schematic illustration of a prior and a posterior distribution*

Once we observe the data, we can make use of Bayes' theorem to find the corresponding 'posterior' probability density. Since some values of the parameters are more consistent with the data than others, we find that the posterior distribution is narrower than the prior distribution. This phenomenon is known as *Bayesian learning.*

It is clear that, since each data point is regarded as simply being some point in a probability distribution, the Bayesian technique models not the data, but the function underlying the data – making it unique in terms of ANN training algorithms. This is an obvious advantage over other training algorithms, but another, less obvious advantage is that, since it models the function, a validation set is not required, allowing the researcher to utilise *all* the data for training (minus a data set for testing).

[3] The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values.

Yet another advantage of Bayesian techniques is that of optimisation. When using other training algorithms, it is necessary to instruct the ANN how accurately you wish it to train. This is called the error goal, and represents the discrepancy between the data passed to it, and its representation of the data. Training to an error goal that is too large may not allow the ANN to best represent the complexity of the data, while training to an error goal that is too small will allow the ANN to learn the noise on the data as well as the data themselves. When using a Bayesian algorithm, since Bayesian techniques are model-driven, training will continue until the optimum model for the data is found, regardless of any error goal, essentially meaning that the optimum error goal for the data need not be passed to the ANN – it will be determined automatically as part of the learning process.

Another advantage of the BR algorithm is its ability to automatically optimise the ANN architecture during the training process. An ANN that is not sufficiently complex (too few neurons) can fail to detect fully the signal in a complicated dataset, leading to underfitting. An ANN that is too complex (too many neurons) may fit the noise, not just the signal, leading to overfitting, and can produce wild predictions even with noise-free data (see Section 3.6.4 for an explanation of the terms underfitting and overfitting). When deciding upon an ANN architecture, it is necessary to take account of the input and output data; the number of input neurons is usually matched to the number of input parameters, and the number of output neurons is matched to the number of output categories required. However, it is not clear how many hidden neurons will be needed for the architecture to be optimum for the required purpose. Many ANN proponents advocate various 'rule-of-thumb' methods for determining the number of hidden neurons, however, different 'rules' lead to differing numbers of neurons in the hidden layer (see Section 3.6.4).

Fortunately, since the BR algorithm is essentially a model-driven (rather than data-driven) algorithm, it allows different models to be compared using only the training data. To gain some insight into this, consider a hypothetical example of three different models, H1, H2 and H3, which we suppose have steadily increasing flexibility, corresponding to a steadily increasing number of hidden neurons. The more complex models can represent a greater range of functions. Figure 3.5d shows a schematic example of the three models, showing the probability (known as the *evidence*) of different data sets D given each model H(i).

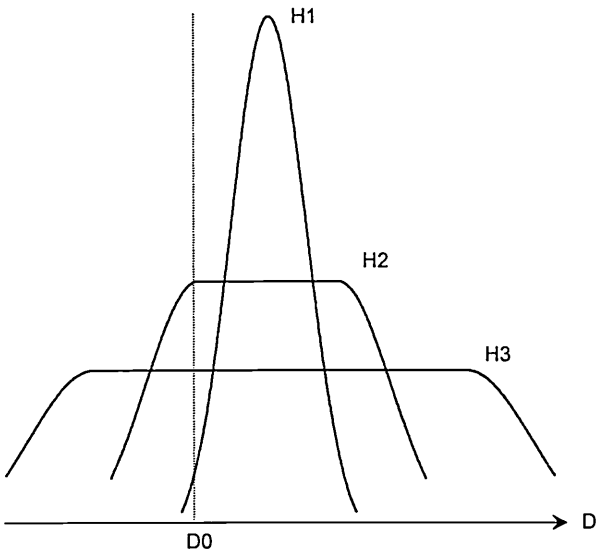



*Figure 3.5d: Schematic example of three models, H1, H2 and H3, which have successively greater complexity*

It can be seen that more complex models can describe a greater range of datasets. Note, however, that the distributions are normalised. Thus, when a particular dataset D0 is observed, the model H2 has greater evidence than either the simpler model H1 or the more complex model H3. In simple terms, the evidence favours models that are neither too simple, nor too complex (MacKay 1992a, Bishop 1995).

This indicates that the Bayesian approach can be used to select a particular model for which the evidence is largest. Practically, this is achieved by the BR algorithm giving an output on training that provides a measure of how many ANN parameters (weights and biases) are being effectively used by the network. This means that a network with too many neurons will be (automatically) pruned to the optimal architecture. This is achieved by the BR algorithm containing a 'penalty term' that penalises weights that are too large, effectively 'switching off' the neurons connected to these ineffective weights.

The MATLAB Neural Net Toolbox version 4.0.1 (Demuth and Beale, 2001) includes an algorithm called Bayesian Regularisation (BR), which is essentially the LM algorithm with a modification to allow the use of Bayesian statistics rather than maximum likelihood statistics in its modelling.

Danaher *et al.* (2004) reported that the BR algorithm produced more accurate ANNs than the LM algorithm when modelling high temperature erosion behaviour of Ni-base alloys. Kumar, Merchant and Desai (2004) reported that the BR algorithm resulted in superior performance compared to the BP algorithm in pulse radar detection.

Summary of Selected Training Regimes

New algorithms are being developed daily to contend with the ever-increasing number of applications for ANNs. However, it is clear that the majority of training algorithms are either supervised or unsupervised, and are feed forward or feed back. Essentially, this gives us four different categories in which to assign the various algorithms. Table 3.5a shows one possible taxonomy for the various algorithms, based on Kosko (1990). It breaks ANNs down in terms of how they are encoded (how they store knowledge) and how they are decoded (how, once trained, they process new input data).

| | | **Decoding** | |
|---|---|---|---|
| | | **Feed forward (FF)** | **Feed back (FB)** |
| **Encoding** | **Supervised** | Delta rule<br>BP<br>BP with momentum<br>LM<br>BR | Recurrent BP |
| | **Unsupervised** | Kohonen self-organising map | Hopfield net |

*Table 3.5a: The taxonomy of algorithms with categorised examples (modified from Kosko 1990).*

Most of the emphasis in ANN applications has been in supervised FF networks; hence, the examples discussed herein are algorithms used in supervised FF-ANNs, and they illustrate the utility (and history) of these particular algorithms. The list is by

no means complete (and the boundaries by no means fixed); indeed the algorithms form only a small percentage of those used in supervised FF-ANNs. Recurrent BP, Kohonen self-organising maps and Hopfield nets are examples of supervised FB, unsupervised FF and unsupervised FB algorithms, respectively. These will not be discussed in further detail here.

## 3.6 Training Issues of Artificial Neural Networks

There are many issues that an ANN investigator needs to be aware of when designing and training an ANN for a specific task. Some of the questions that need to be answered include 'Is my data in an optimum format for training?', 'Will my ANN have the ability to generalise sufficiently well for my requirements?', 'Is the architecture of my ANN optimum for the given problem?', and 'Does my data have noise, and if so, what can I do about it?'. The following sections attempt to give some basic information on how to begin to answer these questions.

It should be noted that the following sections are concerned with FF-MLP ANNs.

### 3.6.1 Data Standardisation

The performance of an ANN is highly dependent upon the data that it is presented with, and the form of the data has a large effect on its perceived quality. It has become standard to perform some pre- and post-processing on the data to ensure that the ANN receives it in the optimum form for training.

#### Pre-processing

The term 'pre-processing' describes any process that converts input data into a form suitable for use with an ANN (DTI 1994). While ANNs can extract essential information from raw input data, the ANNs required can be complex and the computation times very long. A dramatic reduction in both can be achieved by pre-processing the data to derive appropriate inputs. The aims of pre-processing can usually be simplified to three actions:

- **Data selection:** Selecting the most relevant data includes simple operations such as filtering or taking combinations of inputs to optimise the information content of the data. This is particularly important when the data is noisy or contains irrelevant information. Careful selection of relevant data will make the ANN easier to train and improve its performance on noisy data.
- **Data minimisation:** Minimising the number of inputs to the ANN can often simplify a problem greatly. For example, in an image processing application to classify cell types from microscope images, each image having dimensions of 512x512 pixels would contain over a quarter of a million pixels. Clearly it would not be feasible to use so many inputs. In this case, the pre-processing might involve the resizing of the image and/or computing some simple parameters such as length/height ratio of particular regions of the image that could be used as inputs to the ANN.
- **Data transformation:** Transforming (this term is used interchangeably with the terms rescaling and standardising) the data can often simplify the processing that the ANN has to perform and lead to faster training. Although many experienced ANN investigators advocate rescaling the input data to the interval [0,1], this is by no means mandatory. Standardising the input variables tends to make the training process better behaved by ensuring that the random initial weights and biases are of appropriate value. The contribution of an input will depend heavily on its variability relative to other inputs; for instance, if one input has a range of 0-1, while another has a range of 0-1,000,000, then the contribution of the first input will be swamped by that of the second. It is essential, therefore, to rescale the inputs so that their variability reflects their importance.

Post-processing

The term 'post-processing' describes any process that is applied to the output of the ANN. As with pre-processing, it is entirely dependent on the application. If the output transfer function has a range of [0,1], then the target values must lie within that range, however, it is generally better to choose an output transfer function suited to the distribution of the targets than to force the data to conform to the output transfer function.

### 3.6.2 Generalisation

A key attribute of a trained ANN is its ability to meaningfully respond to novel patterns (i.e. to give accurate predictions on unseen data), an ability known as 'generalisation' (Levin *et al.* 1990, Wolpert 1990). For any practical application, it is necessary to ensure that:

- the inputs to the ANN contain sufficient information pertaining to the target, so that there exists a mathematical function relating correct outputs to inputs with the desired degree of accuracy. An ANN cannot be expected to learn a non-existent function (such as a series of random numbers).
- the function to be learnt (that relates inputs to correct outputs) be, in some sense, smooth. In other words, a small change in the inputs should, typically, produce a small change in the outputs. Very non-smooth functions such as those produced by pseudo-random number generators and encryption algorithms cannot be generalised by ANNs.
- the training cases are a sufficiently large and representative subset of the set of all cases that the ANN is required to generalise to. The importance of this condition is related to the fact that there are two different types of generalisation: interpolation and extrapolation. Interpolation applies to cases that are surrounded by nearby training cases. Everything else is extrapolation. In particular, cases that are outside the range of the training data require extrapolation. Cases inside 'holes' in the training data (i.e. under-represented regions of the dataspace) may also effectively require extrapolation. Interpolation can often be done reliably, but extrapolation is notoriously unreliable. Hence it is important to have sufficient training data to avoid the need for extrapolation.

When designing ANNs for a given purpose, it is useful to remember Occam's razor (Chapter 0). In terms of ANNs, this may be interpreted as: do not use an ANN with a large number of neurons when an ANN with a small number of neurons will work just as well. A very useful rule-of-thumb is that for an ANN to be able to generalise, it should have fewer adjustable parameters (weights and biases) than there are data points (Hagan *et al.* 1996, Moody 1992). This technique is termed 'model selection'.

### 3.6.3 Noise

When an ANN makes a prediction it will have an error that may be separated into three terms; model bias[4], model variance[5] and noise (Geman *et al.* 1992). These terms may be further separated into the following six sources (Penny and Roberts 1997):

- **Model bias from data.** This occurs in regions of input space where the ANN output is not equal to the conditional mean of the data. If the new data point falls in a region of high bias the ANN prediction may be highly inaccurate.
- **Model bias from training.** This occurs when an ANN is sufficiently structurally complex to learn the conditional mean of the data in a given region of input space, but the optimisation algorithm does not find the appropriate solution. This can occur if training is stopped too early.
- **Model variance from data.** If an ANN were to be trained on different subsamples of the same dataset, the ANN response would be different in areas of low training data density. Thus, if the test data point is far away from the training data it is likely that the prediction error will be large. This type of prediction error is greatest for complex ANNs.
- **Model variance from training.** An error surface will typically have many local minima, and often more than one global minimum, therefore different randomly chosen initial weights and biases will lead to different minima, and hence, different solutions.
- **Target noise.** If the input variables do not inherently predict the target variable exactly, then there will be some irreducible uncertainty in the predicted value of that target. Target noise may vary non-uniformly over the input space.

---

[4] Bias: A measurement procedure is said to be biased if, on the average, it consistently favours some outcomes over others, giving an answer that differs from the expected value. The bias is the average difference between the measurement and the expected value and can be characterised as a measure of an ANN's ability to generalise correctly to a test set once trained.

[5] Variance: The variance is a measure of the average distance between each of a set of data points and their mean value, and is essentially a measure of the spread of the data. Larger values of variance correspond to data that is more spread out. The variance of an ANN can be characterised as a measure of the extent to which the output of an ANN is sensitive to the data on which it was trained, i.e. the extent to which the same results would have been obtained if a different set of training data were used.

- **Input noise.** This represents intrinsic uncertainty in the values of the input variables, such as experimental measurement error. This leads to uncertainty in the predicted value, as the noise is propagated through the ANN.

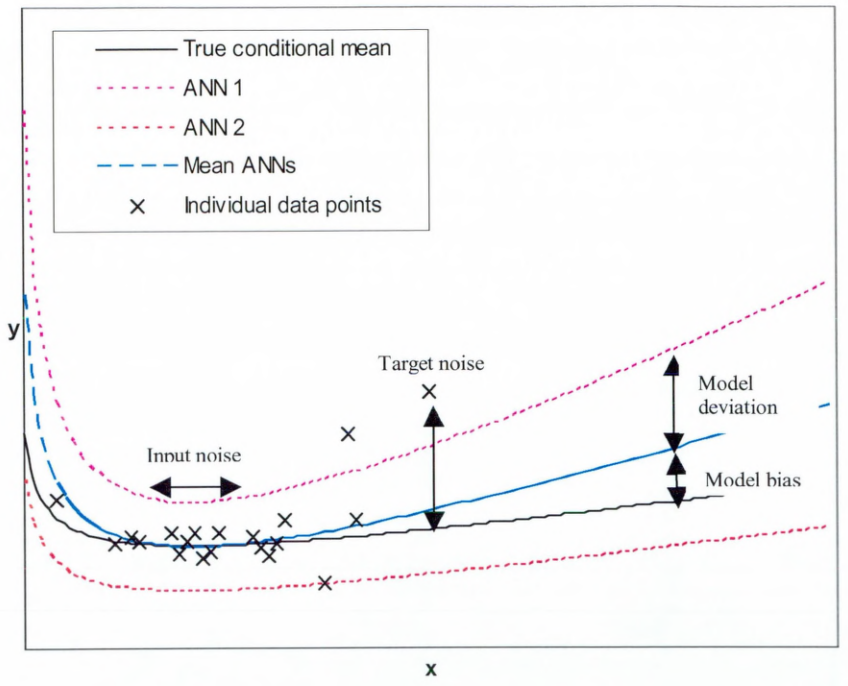



*Figure 3.6a: Separation of the various contributions to the total prediction error in an ANN regression problem. The black line is the true conditional mean of the underlying function of the data to be learnt, the data points are noisy samples of that function. Two independent ANN outputs are shown by the dotted lines and the mean of the two ANN outputs is shown by the dashed line*

Figure 3.6a shows how the various error components may be displayed graphically. It can be seen, for example, that the model is biased for $x > 14$, but not for $x < 14$, has zero variance for $8 < x < 14$, but exhibits high variance outside this range, and the target noise is small for $x = 10$, but large for $x = 15$.

When ANN proponents speak of noise, they are usually referring to measurement error (an example of input noise). For instance, suppose that two samples of soil have identical composition in terms of SSC, %OM and BD. Now also suppose that these samples are found to exhibit slightly different hydraulic properties (differing values of water retention, for example). If both of these samples are presented to the ANN, whether at the training or testing stages, there will clearly be intrinsic limits on the accuracy of generalisation that can be achieved no matter how extensive the training

set is. Usually, the noise distribution is assumed to have zero mean and finite variance, and if this is not the case, the more that is known about the noise distribution, the more effectively the ANN can be trained (McCullagh and Nelder 1989).

### 3.6.4 Underfitting and Overfitting

The critical issue in developing an ANN is generalisation: how well will the ANN make predictions for unseen data? An ANN that is not sufficiently complex can fail to detect fully the signal in a complicated data set, leading to underfitting. An ANN that is too complex may fit the noise, not just the signal, leading to overfitting, and can produce wild predictions even with noise-free data.

Well-known problems with ANN training are the selection of the appropriate topology of the network and overfitting the training data. An extension of the BP algorithm that uses a weight-elimination technique (Weigand *et al.* 1990) addresses both problems. The idea is to start with too many hidden neurons and to introduce into the training algorithm a term that penalizes large weights on the connections between neurons. Thus, the algorithm, during training, eliminates an appropriate number of weights and neurons in order to obtain the appropriate generalisation on the training data, thereby attaining the optimum topology that introduces the necessary model complexity without overfitting.

### 3.6.5 Early Stopping

If a validation set is used during training, the errors on the training and validation sets may be monitored as the ANN training proceeds. During training, errors on the training set will usually decrease; however, this is not always the case with errors on the validation set. When errors on the validation set start to increase, the ANN is starting to model the noise in the data as well as the signal (overfitting). At this point, the ANN, although trained accurately on seen data, will not produce accurate predictions on unseen data, and the training should be stopped. This method of ending the training is referred to as 'early stopping'.

## 3.7 Testing Artificial Neural Networks

When testing ANNs, what is being tested is the ability of the ANN to generalise, and the outcome of the testing procedure is often termed the 'generalisation error'. There are many methods of estimating the generalisation error of an ANN, and the most common – split-sample validation, cross-validation and Bootstrapping – are briefly discussed here.

### 3.7.1 Split-Sample Validation

The most commonly used method for estimating generalisation error in ANNs is to reserve part of the data to be used as a test set, which is not used during the training process, and must be a representative sample of the cases that the ANN is to generalise to. After training, the test data is presented to the ANN, and the error on the test set provides an unbiased estimate of the generalisation error, provided that the test set was chosen randomly. The disadvantage of split-sample validation is that it reduces the amount of data available for both training and validation, although this is not a problem if data is in abundance.

### 3.7.2 Cross-Validation

In cross-validation, (also known as k-fold cross-validation) the data is divided into k subsets of equal size. The ANN is trained k times, each time leaving out one of the subsets from training, but using only the omitted subset to compute the generalisation error of that training subset. If k equals the sample size, this is called 'leave-one-out' cross-validation.

The distinction between cross-validation and split-sample validation is extremely important because cross-validation has been shown to be markedly superior for small data sets (Goutte 1997), and allows all of the data to be used for training. The disadvantage of cross-validation is that the ANN has to be retrained many times.

### 3.7.3 Bootstrapping

In the simplest form of Bootstrapping, instead of repeatedly analysing subsets of the data, random subsamples of the data are sampled (and replaced, so that the original dataset remains unchanged) and analysed. Bootstrapping is an improvement on cross-validation that often provides better estimates of generalisation error at the cost of increased computing time (Efron 1983).

## Chapter Conclusions

ANNs have advantages over many other modelling methods in that they can 'learn' by example without requiring information *a priori*. This is achieved by feeding a signal forwards through layers of (comparatively) simple, interconnected neurons, so that a prediction of the target function may be seen at the output stage. When the training is supervised, this prediction is compared to the expected target function, and the difference is fed backwards through the layers to allow the ANN to 'learn' from the difference, and ultimately produce improved estimates of the target function.

In terms of training algorithms, the LM algorithm is widely acknowledged to provide the fastest convergence on function approximation problems for ANNs that contain up to a few hundred weights. However, the BR algorithm, which is essentially the LM algorithm with a modification to allow the use of Bayesian statistics, is unique in that it is model-driven (rather than data-driven) and allows the effective modelling of the *function* underlying the data. Other advantages of using the BR algorithm over competing algorithms are that an ANN does not require a validation set, and both the error goal and the ANN architecture are determined automatically.

# Chapter 4

# Combining Artificial Neural Networks

This chapter introduces the previous research that has been conducted into using artificial neural networks for pedotransfer function construction, and discusses some of the drawbacks of current methods outlined in the literature. It also briefly introduces the concept of splitting a large dataset into a number of smaller, more manageable datasets, and explains why this may be beneficial to artificial neural network models. Splitting data into smaller sub-sections of data is not difficult, provided that the integrity of the input-output function is not compromised. In simple terms, this means that we should take care with data selection, and this is discussed here in theoretical terms. The difficulty is in knowing exactly how to combine the micro-models once they have been built, and the various methods of achieving this are outlined here.
This chapter details the two main modes of combining ANNs; ensemble and modular methods, along with the critically important concept of the bias/variance trade-off.

## 4.1 Artificial Neural Networks For Pedotransfer Functions

The benefit of developing PTFs through an ANN is that it is not necessary to formulate an *a priori* mathematical model for the prediction, and this approach is ideal for utilising non-uniform input data.

The use of ANNs to develop PTFs is not new. A BP-MLP has been used almost exclusively to develop PTFs by researchers such as Pachepsky *et al.* (1996), Schaap and Bouten (1996), Schaap *et al.* (1998), Koekkoek and Booltink (1999) and Minasny *et al.* (1999). The BP-FF radial basis function (RBF) ANN used by Tamari *et al.* (1996) for prediction of $K_{sat}$ represents an exception. The RBF-ANN has advantages over the MLP-ANN in that it is quicker and has a greater non-linear capability, although it has limited interpolation capabilities and is not suitable for problems with many inputs (the number of hidden units grows exponentially with the number of inputs).

The overall conclusion made by these investigators was that when the number of inputs is greater than three, ANNs usually perform better than regression techniques, particularly when uncertainties in the quality of the data were small. Grouping samples by horizon also improved the precision of the estimates, especially in subsoil, since the subsoil data is less variable.

It is for this reason that I have chosen ANNs as my medium of investigation with which to construct continuous and class PTFs (see Chapters 6 to 8).

Uses of ANNs by previous researchers have been consistent in one respect – they have all trained a single ANN to model the input-output dependency of the data (with the exception of Schaap and Leij (1998a), Schaap and Leij (1998b), Schaap *et al.* (1998), Schaap *et al.* (2001) and Nemes *et al.* (2003). These papers will be discussed in further detail in Section 6.3.1). Typically, this means taking the data from perhaps 2000 soil samples, using roughly 2/3 of these in the training regime and the remaining 1/3 for testing. Thus, a very large data set is used to build the model, taking a long time to adequately train an ANN of sufficient complexity. It is then assumed that the data used for training contains the relevant information to accurately and reliably represent the underlying function of the data. These assumptions need to be challenged.

The ANNs used by previous researchers tend to err towards more complex models, however, Occam's razor suggests that we should prefer simpler models. This project has focussed on the task of increased simplicity of model construction. Of course, the data presented to a model should possess the necessary complexity to fully describe the underlying function, and thus, this project attempts to balance the requirements of data complexity and model simplicity.

My research is focussed towards the concept of splitting a large and complex soil dataset into a number of 'bite-sized' chunks, with the aim of building 'micro-models', which will then be combined in some way at a later stage to form the main 'macro-model'.

## 4.2 Artificial Neural Network Ensembles

In general, different ANN models provide different generalisations. It is therefore useful to train a number of ANNs to ensure that a good model/parameter set is found.

However, selecting the 'best' ANN is not necessarily the ideal choice, because potentially valuable information may be wasted by discarding the results of less-successful individual ANN models (Perrone and Cooper 1993, Tumer and Ghosh 1996). This leads to the concept of 'combining', where the outputs of several classifiers are pooled before a decision is made (Tumer and Ghosh 1996).

The expression 'ANN ensemble[1]' (also known as 'committee' or 'committee machine') is commonly used for the combining of a set of individual 'redundant' ANNs. The component ANNs are redundant in the sense that each provides a solution to the same task, even though this solution may be obtained by different means.

The main reasons for combining ANNs in redundant ensembles are to improve the ability to generalise and to guard against the 'failure' of individual component ANNs. Here, the term fail refers to the fact that ANNs will have been trained on a limited dataset and, on the basis of that data, asked to estimate the target function. These estimates will not usually be identical to the target function, and will usually under- or over-estimate the expected value(s).

Ensembles have a long history in the real world, for instance:

- The Condorcet Jury model proposed in 1786 that a democracy as a whole is more effective than any of its constituent members (Grofman and Owen 1986).
- In forecasting, it has been shown that better results can be achieved by combining forecasts than by choosing the best one (Bates and Granger 1969).

It has been shown (Perrone and Cooper 1993) that the mean squared error (MSE) of an ensemble is related to the MSE of its constituent members by:

$$MSE_{Ensemble} = \frac{1}{N}\overline{MSE} \tag{4.2a}$$

where $\overline{MSE}$ is the average MSE of the individual ANNs and $N$ is the number of member ANNs in the ensemble. This implies that by increasing the ANN population size, the error of an ensemble's power of estimation can be made to be arbitrarily

[1] Ensemble is a French word, meaning 'together' or 'at the same time', and usually refers to a unit or group of complementary parts that contribute to a single effect.

However, selecting the 'best' ANN is not necessarily the ideal choice, because potentially valuable information may be wasted by discarding the results of less-successful individual ANN models (Perrone and Cooper 1993, Tumer and Ghosh 1996). This leads to the concept of 'combining', where the outputs of several classifiers are pooled before a decision is made (Tumer and Ghosh 1996).

The expression 'ANN ensemble[1]' (also known as 'committee' or 'committee machine') is commonly used for the combining of a set of individual 'redundant' ANNs. The component ANNs are redundant in the sense that each provides a solution to the same task, even though this solution may be obtained by different means.

The main reasons for combining ANNs in redundant ensembles are to improve the ability to generalise and to guard against the 'failure' of individual component ANNs. Here, the term fail refers to the fact that ANNs will have been trained on a limited dataset and, on the basis of that data, asked to estimate the target function. These estimates will not usually be identical to the target function, and will usually under- or over-estimate the expected value(s).

Ensembles have a long history in the real world, for instance:

- The Condorcet Jury model proposed in 1786 that a democracy as a whole is more effective than any of its constituent members (Grofman and Owen 1986).
- In forecasting, it has been shown that better results can be achieved by combining forecasts than by choosing the best one (Bates and Granger 1969).

It has been shown (Perrone and Cooper 1993) that the mean squared error (MSE) of an ensemble is related to the MSE of its constituent members by:

$$MSE_{Ensemble} = \frac{1}{N}\overline{MSE} \tag{4.2a}$$

where $\overline{MSE}$ is the average MSE of the individual ANNs and $N$ is the number of member ANNs in the ensemble. This implies that by increasing the ANN population size, the error of an ensemble's power of estimation can be made to be arbitrarily

[1] Ensemble is a French word, meaning 'together' or 'at the same time', and usually refers to a unit or group of complementary parts that contribute to a single effect.

small when compared to the average error of the ANNs when taken as individuals. In practice, as $N$ becomes large, only small improvements, if any, may be made to the predictive ability of the ensemble. This is due to certain assumptions about the data, the details of which are beyond the scope of this thesis. Further details may be found in Perrone and Cooper (1993).

Equation 4.2a is a very powerful result, not only because of its effect on ANN estimators, but more crucially, since equation 4.2a does not specifically mention ANNs, the method of combining estimators in ensembles holds for any type of estimator, providing that an error estimate is given. Indeed, Mukkamala, Sung and Abraham (2004) showed that an ensemble of ANNs, SVMs (Support Vector Machines) and MARS (Multivariate Adaptive Regression Splines) produced superior results to any of these three methods when used individually, when applied to intrusion detection for the protection of information systems security.

### 4.2.1 The Bias/Variance Trade-Off

Combining a set of imperfect predictors (see Section 4.2.3 for more details on this) is a way of managing the limitations in the individual predictors; that is, each component ANN is known to make errors (bias, variance and noise, see Section 3.6.3), but they are combined in such a way as to minimise the effect of these errors. The effect of combining ANNs to reduce errors may be expressed in terms of the statistical terms bias and variance.

It has been shown that the MSE of an individual predictor can be expressed in terms of the variance and the bias (for example, see Raviv and Intrator 1996, Sharkey 1999, Wang 1998). If $\hat{\theta}$ is an estimator of the quantity $\theta$, then the mean squared error can be expressed as:

$$MSE = Variance\left(\hat{\theta}\right) + Bias^2\left(\hat{\theta}\right) \tag{4.2b}$$

Figure 4.2a is an illustration of equation 4.2b, and shows that the model bias is a decreasing function, whilst the model variance is an increasing function, of the model complexity. This seems intuitively correct, since, as the complexity of the model is

increased (i.e. increasing the amount of data), the variety of the data samples in the model is increased, and so the bias towards individual data samples must decrease.

Figure 4.2a shows that there is a trade-off between bias and variance in terms of training ANNs; the best generalisation (strong performance on unseen data as well as on seen data) requires a compromise between the conflicting requirements of small variance and small bias to produce an optimal fitting. It is a trade-off because attempts to decrease the bias (by adding data samples to the model from under-represented regions of the dataspace) are likely to result in higher variance, and *vice versa*. An ANN that is set up to generalise well (an optimal compromise between low variance and low bias) must take sufficient account of the data to avoid bias towards a particular sub-set of the data, but avoid overfitting (Sharkey 1999).

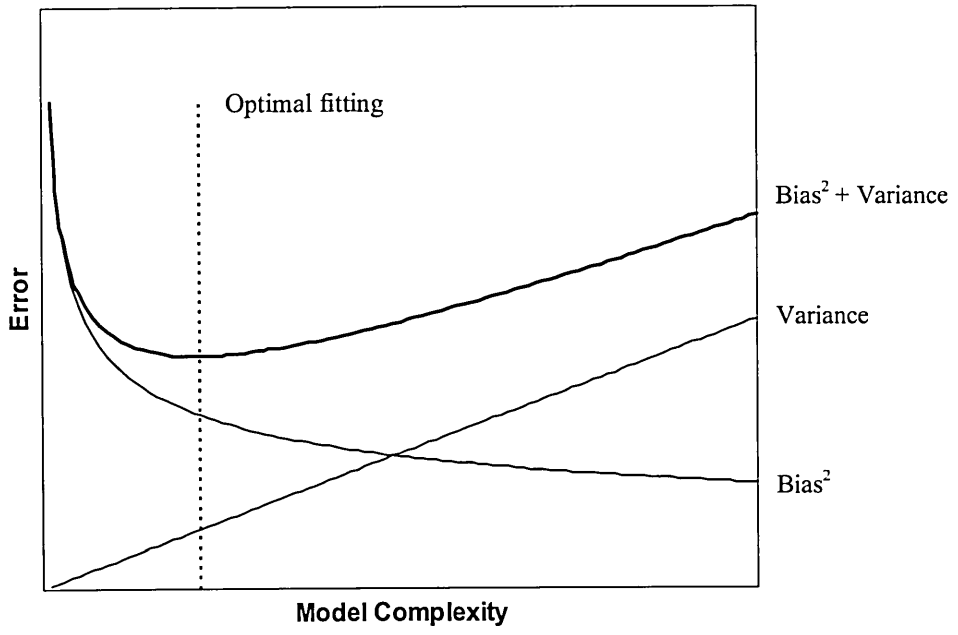



*Figure 4.2a: Illustration of the bias/variance trade-off for a single estimator*

In terms of an ensemble of ANNs, the bias measures the extent to which the ensemble output averaged over all the ensemble members differs from the target function, whilst the variance is a measure of the extent to which the ensemble members disagree (Krogh and Vedelsby 1995).

The improvement of predictions that can arise from ensemble combinations is usually the result of a reduction (if not complete elimination) in variance, whilst leaving the bias unaltered (for example, see Tumer and Ghosh 1996, Wang 1998). Therefore, an effective approach is to create a set of ANNs that exhibit high variance, but low bias, since the variance component can be removed by combining. In other words, one should take more account of the available data to reduce bias, since the increase in variance will be removed by combining individual predictions. Ensembles can thus provide a way of reducing overall prediction errors by *using* the bias/variance trade-off, rather than being *restricted* by it (Sharkey 1999).

In ensembles having a large number of ANN members, Sollich and Krogh (1996) found that individual ANNs should be over-trained in order to maximise the benefits of the variance-reducing effects of ensemble learning. For ensembles of more realistic size, they found that optimising the ANNs creates an ensemble that yields better generalisation performance than a single ANN.

### 4.2.2 Creating Ensemble Members

Since the main reason for combining ANNs in ensembles is to improve predictive ability, there is clearly no advantage to be gained from an ensemble that is composed of a set of identical ANNs (since the bias of the individual ANNs will be identical, and the variance of the individual predictions equal to zero). Thus, the aim is to create ANNs that generalise differently. There are basically four parameters that may be changed to achieve this. A set of ANNs can be created by varying the:

- **Initial conditions** (initial random weights and biases from which each ANN is trained).
- **Topology** (architecture of each ANN – training with a varying number of hidden units).
- **Training Algorithm.**
- **Training Data** (or the form of the training data).

The fourth of these techniques requires a little more clarification. In Section 4.1 I introduced the concept of splitting a large and complex soil dataset into a number of 'bite-sized' chunks. To illustrate this, consider a dataset $H$ that may be divided into

segments *A* to *E* according to some difference in functional behaviour, Figure 4.2b, and let each segment contain 100 data samples.

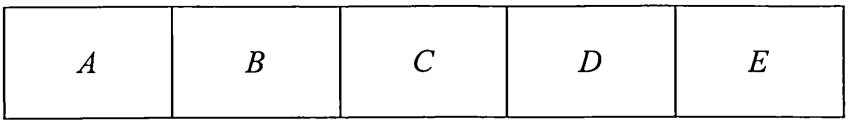



*Figure 4.2b: Dataset* H, *divided into segments* A *to* E

Let us suppose that we require 50 data samples in an ANN modelling process. Extracting all 50 data samples at random from segment *A* and training a single ANN with this data would lead to a model that has low variance and a strong bias towards segment *A*. When testing with unseen data, the ANN would produce good predictions for segment *A*, but poor predictions for segments *B* to *E*. This method of modelling is evidently undesirable. An alternative method would be to extract ten samples randomly from each of the five segments. Training a single ANN with this data would produce a model that has high variance and generates unbiased predictions across the five segments when tested with unseen data. Let us now consider repeating this alternative method a number of times to produce an ensemble of ANNs. Each ANNs training dataset is low in bias and high in variance. The individual ANNs would produce equally accurate predictions (on the average) across the five segments when tested with unseen data. However, the error on each prediction has both a bias and a variance component. When taking the average of the individual ANN predictions to form an ensemble prediction, the variance component of the error is eliminated, resulting in an improvement in the prediction. Thus, by splitting the dataset into a number of regions and extracting data randomly from each of these regions leads to ANNs that are high in variance and low in bias – precisely the conditions we are seeking in order to best exploit the bias/variance trade-off.

These four techniques of creating ensemble members outlined above may be employed individually, or in combinations of two or more to increase variance in the ensemble model. In the continuous ANN-PTF ensemble model of water retention (see Section 6.2.2), I will demonstrate how I increase variance in the model by varying the training data and the initial conditions of each individual ANN.

### 4.2.3 Creating Ensembles (Combining Members)

One striking aspect of combining ANNs is the manner in which local minima on the error surface are dealt with. Most ANN algorithms achieve sub-optimal performance specifically due to the existence of large numbers of sub-optimal local minima. Combining a set of ANNs that have converged to local minima leads to improved estimates. In general, ANNs that have converged to local minima will perform poorly in different regions of feature space and thus their error terms will not be strongly correlated. It is this lack of correlation that drives the combining method, for whichever method is employed. Thus, combining ANNs efficiently utilises the local minima that other techniques (including single ANN methods) try to avoid (Perrone and Cooper 1993).

Several methods exist for combining the outputs of ANN members, and some of the more common methods are briefly mentioned here:

- **Averaging:** Taking the statistical mean of the outputs is perhaps the simplest of all methods (Tumer and Ghosh 1995).
- **Weighted averaging:** Taking the weighted average of the outputs takes account of the relative accuracies of the ANNs to be combined (Tumer and Ghosh 1995, Wolpert 1992).
- **Non-linear combining methods:** Non-linear methods include belief-based methods (Kolen and Pollack 1990), combining using rank-based information (Al-Ghoneim and Vijaya Kumar 1995), voting and weighted voting (Hansen and Salamon 1990), and order statistics (Tumer and Ghosh 1995).
- **Stacked generalisation:** Stacked generalisation is a scheme for minimising generalisation error by introducing a second level of generalisers whose inputs come from the predictions of a first level of units after being trained with part of the learning set and generalising on the rest (Wolpert 1992).
- **Supra Bayesian:** The principle underlying the Supra Bayesian approach is that from the viewpoint of the decision maker, the opinions expressed by experts (i.e. the predictions of the ANNs) are themselves data. Consequently, a decision maker combines the probability distributions provided by the individual experts via Bayes' rule (Jacobs 1995).

Perrone and Cooper (1993) showed that an ANN ensemble whose individual predictions are combined by the weighted averaging method always generates predictions that are as accurate as, or more accurate than, the best individual ANN or ANN ensembles whose individual predictions are combined by the averaging method.

Since Occam's razor is a key feature of my research, the continuous ANN-PTF ensemble model of water retention (see Section 6.2.2) will utilise the weighted averaging method, which is superior to the averaging method, but simpler than the other methods outlined above.

## 4.3 Modular Artificial Neural Networks

Modularity can be defined as the subdivision of a complex object into simpler objects, and examples are to found in all areas of nature. At the atomic level, electrons, protons, and neutrons make the building blocks for atoms. Atoms form elements and various elements form simple and more complex molecules, of which everything is constructed. In living creatures, proteins form cells, and at a higher level cells could be seen as basic components. This idea of modules can be continued to more and more complex structures, even looking outwards at the universe – planets can be seen as modules within a solar system, and solar systems as modules in galaxies, etc..

Replication and decomposition are the two main concepts for modularity. Replication is a way of reusing knowledge. Once one module is developed and has proved to be useful it is replicated in larger numbers. This principle can be seen in living organisms such as the human, which has two legs, eight fingers (and two thumbs), several vertebrae, thousands of hair follicles, and billions of cells. Decomposition is often found when dealing with a complex task. It is common to solve a complex problem by decomposing it into simpler tasks that are easier to manage and then reassemble the solution from the results of the subtasks. In other words, a task could be solved with a single ANN, but better performance is achieved when it is broken down into a number of specialist modules.

The concept of modules differs from the concept of the ensemble in that the ensemble is made up of a number of redundant subtasks where the complete task solution is a decision fusion of the individual task solutions, Figure 4.3a (left).

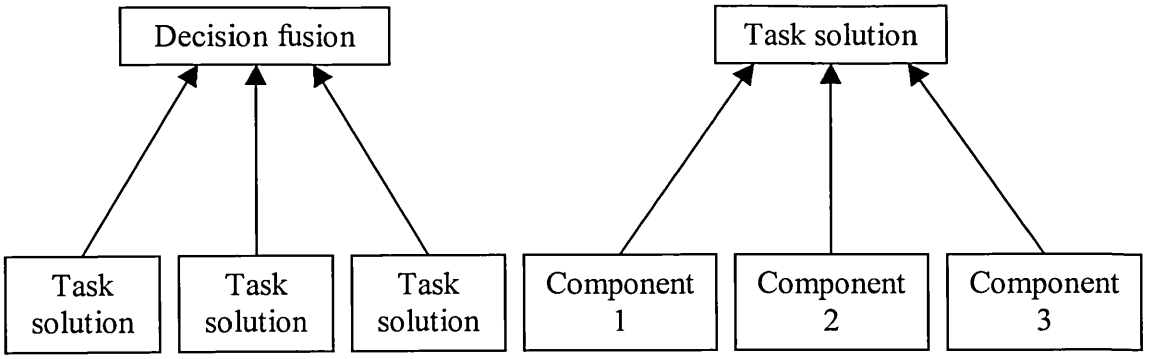



*Figure 4.3a: Illustration of the difference between ensembles (left) and modules (right)*

Modules and ensembles may be combined in any way that is deemed necessary. For instance, a large dataset may be subdivided into a number of smaller, more manageable subsets, and each of these subsets may then be considered as a module, since they form a specialised subset of the whole problem. Each individual subset may then be modelled using ensembles, so that each module is also an ensemble.

Each component in a modular system can take the form of an ANN; however, as in ensemble combinations, there is no reason why some of the components could not make use of non-neural computing techniques. For instance, one particular subset of the data may be modelled with an ANN ensemble, whilst another may be modelled by multiple linear regression.

### 4.3.1 Creating Modular Components

The decomposition of a problem into modular components may be accomplished explicitly, by class decomposition or by automatic means.

- **Explicit decomposition:** Before learning, a problem is divided into a set of sub-problems by the designer who has some domain knowledge and substantial prior knowledge concerning the decomposition of the problem (Lu and Ito 1999). The limitation of this method is the fact that sufficient prior knowledge of the problem is necessary.
- **Class decomposition:** Before learning, a problem is broken down into a set of sub-problems according to the inherent relations among training data (Anand *et*

*al.* 1995). In contrast to the explicit decomposition, this method requires only some common knowledge concerning the class relations among training data.

- **Automatic decomposition:** A problem is decomposed into a set of sub-problems with the progression of learning (Lu and Ito 1999), and is also known as blind data partitioning. From the point of view of computational complexity, explicit and class decomposition techniques are more efficient than automatic decomposition because a problem has been decomposed into sub-problems before learning. The advantage of this method is that it is more general because it can work when prior knowledge is absent.

### 4.3.2 Combining Modular Components

There are three different modes of combining modular component ANNs, by competitive, sequential or supervisory means.

- **Competitive:** In some modular combinations, the best performance is achieved by selecting the most appropriate module depending on the particular circumstances (Sharkey 1999). There are two main mechanisms for this selection, Gating and Rule-based switching. With Gating, the system learns to allocate examples to the most appropriate module, whilst with Rule-based switching, the switching is achieved by a more explicit mechanism.
- **Sequential:** In sequential combination, processing is successive – the outputs of an earlier ANN form the inputs of another (Sharkey 1999).
- **Supervisory:** In a supervisory relationship, one module is used to supervise the performance of another, and perhaps select its parameters on the basis of observations of the effect of various parameter values on the performance of that ANN (McCormack 1997).

## Chapter Conclusions

The benefit of developing PTFs through an ANN is that it is not necessary to formulate an *a priori* mathematical model for the prediction, and this approach is ideal for utilising non-uniform input data. Overall, it has been shown that when the

number of inputs is greater than three, ANN-PTFs usually perform better than regression techniques, particularly when uncertainties in the quality of the data were small. Grouping samples by horizon also improves the precision of the estimates.

An ANN ensemble is typically a collection of individual ANNs that each provide a solution to the same task, and are combined in some way to provide improved estimates of the task solution. The main reasons for combining ANNs in ensembles are to improve the ability to generalise and to guard against the failure of individual component ANNs.

It can be shown that the errors of individual ANNs may be expressed in terms of bias and variance, and there is a trade-off between these terms, such that efforts to decrease bias will result in a concomitant increase in variance, and *vice versa*. The improvement of predictions that can arise from ensemble combinations is usually the result of a reduction (if not complete elimination) in variance, whilst leaving the bias unaltered. Therefore, an effective ANN ensemble is composed of a number of individual ANNs that each exhibit high variance, but low bias. This is achieved by taking more account of the available data to reduce bias, since the increase in variance will be removed by combining individual predictions. Ensembles thus provide a way of reducing overall prediction errors by *using* the bias/variance trade-off, rather than being *restricted* by it.

In contrast to the ensemble method, modular ANNs are specialised ANNs that model a selected sub-section of a complex dataspace, in other words, better performance is achieved when the problem domain is broken down into a number of specialist modules. Generally speaking, it would be expected that modular ANNs would produce improved models over the ensemble or single ANN methods, since modular ANNs are specialised for their given task.

# Chapter 5

# Selection and Statistical Analyses of the Soil Data

Typically, modellers have used various input parameters when constructing ANN-PTFs, such as soil texture, dry bulk density and organic matter content, often in different combinations. These are used to train ANN-PTFs to enable them to make predictions of water retention or hydraulic conductivity, usually at field capacity or permanent wilting point. In this chapter I will justify why I have used SSC, BD and %OM as the input parameters, and water retention at field capacity as the output parameter. I will also apply various selection criteria to the available data, and analyse them using a range of statistical tests to determine the statistical properties of the different parameters and texture classes. The results of these tests will provide a point of comparison for the ensemble models that will be discussed in some detail in Chapters 6 and 7.

## 5.1 Data Selection and Selection Criteria

While some practitioners have preferred to construct parameter prediction ANN PTFs of a hydraulic characteristic (see Section 2.2.3), others have modelled specific points on the WRC or HCC (see Section 2.2.5). The various researchers have often used different levels of input and output with which to build their models, often with quite differing results. Table 5.1a is adapted from Wösten *et al.* (2001), and presents a sample of water retention PTFs from literature.

In these examples, the RMSE of volumetric water contents ranges from 0.02 to $0.11 m^3 m^{-3}$. It is interesting to note the accuracy of the result of Bruand *et al.* (1996), which appears to show that bulk density is the only parameter required to achieve accurate results. This is misleading, and is a consequence of the fact that the soil samples used were high in organic matter. Bloemen (1980) suggested that, as input parameters, organic matter and bulk density are interchangeable, and this indicates that the soil structure in this case is highly dependent upon the bulk density. Accuracy of predicting a complete characteristic is lower than accuracy of predicting a specific pressure head (Wösten *et al.* 2001).

| Source | %Cl | %Si | %Sa | %OM | BD | -h /kPa | RMSE ($m^3m^{-3}$) | Other PTF input variables/notes |
|---|---|---|---|---|---|---|---|---|
| Ahuja *et al.* (1985) | ✓ | ✓ | | ✓ | ✓ | 33<br>1500 | 0.05<br>0.05 | |
| Beke and McCormic (1985) | ✓ | ✓ | | ✓ | ✓ | 33<br>1500 | 0.05<br>0.04 | |
| Bell (1993) | | | ✓ | ✓ | | 1500 | 0.04 | |
| Bruand *et al.* (1996) | | | | | ✓ | 33<br>1500 | 0.03<br>0.03 | |
| Calhoun *et al.* (1972) | ✓ | | | | | 1500 | 0.05 | |
| Gupta and Larson (1979) | ✓ | ✓ | | ✓ | ✓ | 1500 | 0.05 | |
| Koekkoek and Booltink (1999) | ✓ | ✓ | ✓ | ✓ | ✓ | 10<br>1500 | 0.05<br>0.05 | |
| Lenhardt (1984) | ✓ | | | | | 33<br>1500 | 0.07<br>0.05 | |
| Minasny *et al.* (1999) | ✓ | ✓ | ✓ | | ✓ | 33<br>1500 | 0.07<br>0.07 | Porosity, mean particle diameter, geometric standard deviation |
| Pachepsky *et al.* (1996) | ✓ | ✓ | ✓ | | ✓ | 33<br>1500 | 0.02<br>0.02 | |
| Paydar and Cresswell (1996) | | | | | | A | 0.02 | Slope of the particle-size distribution + one measured point on the WRC |
| | | | | | | A | 0.04 | Various textural classes |
| Paydar and Cresswell (1996) | ✓ | ✓ | ✓ | ✓ | | A | 0.03 | Coarse sand and fine sand |
| Schaap *et al.* (1998) | ✓ | ✓ | ✓ | | | A | 0.11 | Texture class only |
| | ✓ | ✓ | ✓ | | ✓ | | 0.09 | |
| Schaap and Leij (1998a) | ✓ | ✓ | ✓ | | | A | 0.10 | |
| Sinowsky *et al.* (1997) | ✓ | ✓ | ✓ | | ✓ | 30<br>1500 | 0.04<br>0.04 | Porosity, median particle diameter, standard deviation |
| Tomasella and Hodnett (1998) | ✓ | ✓ | ✓ | | | A | 0.06 | |

*Table 5.1a: Typical examples of the water retention PTF accuracy (where A represents the average RMSE along the measured water retention curves)*

It can be seen that SSC, BD and %OM, and the (average) inflexion points of the WRC (-33 and -1500kPa) are the most commonly used input and output parameters, respectively. It seems reasonable, therefore, that these should be considered as a useful starting point for modelling. Hence, SSC, BD and %OM are chosen to be the initial input parameters for my ANN-PTF models.

Unfortunately, choice of output parameter is a little more complicated. Most researchers have chosen to model FC water retention at -33kPa; however, the closest value to this in HYPRES is the water retention value at -250cm$H_2O$, which is equal to -24.5kPa. Since the many definitions of FC are very vague (indeed, one definition states that the FC resides at a point between -10kPa and -33kPa on the WRC), I decided that, to simplify matters (Occam's razor), no correction should be applied to the water retention values at -250cm$H_2O$ to force them to apply to water retention values at -33kPa. Thus, the output parameter chosen for modelling is the water retention data held in HYPRES at -250cm$H_2O$, and this is assumed to be the FC, denoted $\theta_{FC}$.

In addition to this, since there are very few organic soils in HYPRES (i.e. those that have %OM > 10) only mineral soils will be considered, as discussed briefly in Section 2.1. Not only are there very few soils in HYPRES for which %OM > 10, but early results indicate that when included in water retention models they yield predictions that are significantly different from what is expected.

| %Cl | %Si | %Sa | BD | %OM | $\theta_{FC}$ | Ensemble Prediction |
|---|---|---|---|---|---|---|
| 0 | 2.54 | 97.5 | 0.934 | 19.7 | 0.551 | 1.773 |
| 0 | 2.2 | 97.8 | 1.001 | 18.2 | 0.476 | 1.752 |
| 0 | 0.67 | 99.3 | 0.979 | 16.1 | 0.526 | 1.794 |
| 10.4 | 22.6 | 67 | 1.185 | 10.3 | 0.470 | 1.397 |
| 41 | 46.8 | 12.2 | 0.84 | 27.3 | 0.635 | 1.205 |
| 42 | 45.5 | 12.5 | 0.82 | 27.3 | 0.665 | 1.232 |
| 49.6 | 41.8 | 8.6 | 0.752 | 24 | 0.651 | 0.923 |
| 74 | 23 | 3 | 0.64 | 19 | 0.652 | 0.926 |

*Table 5.1b: Selected results indicating the unsuitability of using organic soil horizons in models of soil water retention*

Table 5.1b shows selected results for organic soil horizons, and shows that the difference between what is expected ($\theta_{FC}$) and what is predicted by an ensemble of

ANN-PTFs is very large (recall that $\theta_{FC}$ should be in the range [0, 1]). Additionally, the spread of predictions of the individual members is also very large (these results are not detailed here), indicating that the ANNs are not in agreement as to what the prediction should be, and thus confidence in these predictions is very low.

Also, duplicate horizons should be discarded to reduce noise in the model. To explain this, I would like to consider the following hypothetical soil horizons, with their corresponding FC water retention values:

| | Inputs | | | | | Output |
|---|---|---|---|---|---|---|
| | %Sa | %Si | %Cl | BD | %OM | $\theta_{FC}$ |
| **Horizon 1:** | 40.0 | 40.0 | 20.0 | 1.50 | 5.0 | 0.40 |
| **Horizon 2:** | 40.0 | 40.0 | 20.0 | 1.50 | 5.0 | 0.50 |

Soil horizons 1 and 2 are physically identical, but differ in their hydraulic characteristics. This difference could be due to many factors, from differing measurement techniques, to errors in measurement, calculation or insertion into the database. The modeller has no knowledge of the source of the discrepancy, and no way of determining which of the $\theta_{FC}$ values is correct, if indeed either one of them is. It is common knowledge to the soil scientist that the WRC suffers from hysteresis (the curves for loading and unloading differ). Bell and van Keulen (1996) reported that caution is required in using PTFs for FC developed from disturbed soil samples as they over-estimated *in-situ* FC for all soils except for the coarser textured soils. HYPRES contains data that indicates whether samples have been disturbed, indeed separate PTFs could be constructed, one with disturbed and one with undisturbed samples, in order to verify the work of these researchers. This investigation has not been conducted in the research reported in this thesis. The discrepancy between the two possible values of $\theta_{FC}$ represents a source of irreducible error to the modeller, since a model cannot reach a level of zero error on one horizon without yielding a large error on the other. The best that a model can do is to model both outputs at the value of 0.45, essentially giving half of the error to each of the horizons, whilst leaving the overall error unaltered. It is for this reason that I decided that horizon duplicates must be deleted before modelling, to reduce the level of target noise (see Section 3.5.3).

In summary, the parameters to be used in the modelling process (including their criteria) are:

| **Input Parameters:** | **Output Parameter:** |
|---|---|
| %Sa<br>%Si<br>%Cl<br>BD<br>%OM | $\theta_{FC}$ |
| **Selection Criteria:**<br><br>All soil horizon duplicates deleted<br>Mineral soils only (%OM < 10) | |

When these criteria have been applied to, and the data extracted from HYPRES, the data detailed in Table 5.1c resulted.

| | Texture Class | | | | | |
|---|---|---|---|---|---|---|
| | **C** | **M** | **MF** | **F** | **VF** | **ALL** |
| Total number of soil horizons selected | 670 | 1000 | 603 | 587 | 130 | 2990 |
| Number deleted where %OM > 10 | 17 | 6 | 1 | 22 | 7 | 53 |
| Number of deleted duplicates | 23 | 42 | 28 | 49 | 16 | 158 |
| **Total number remaining** | **630** | **952** | **574** | **516** | **107** | **2779** |
| Percentage of total remaining | 23 | 34 | 21 | 18 | 4 | 100 |

*Table 5.1c: Textural breakdown of the data selected for modelling*

Of the 2990 soil horizons selected prior to application of the selection criteria, 211 of these were deleted (53 had %OM > 10, and 158 were duplicates), leaving 2779 viable soil horizons for the modelling process.

Approximately one-third and one-quarter of the selected soil horizons belong to the M and C texture classes, respectively, whilst just over and just under one-fifth belong to the MF and F texture classes, respectively. The remaining 4% belong to the little-represented VF texture class. The distribution of these datapoints on the soil textural triangle are shown in Figure 5.1d.

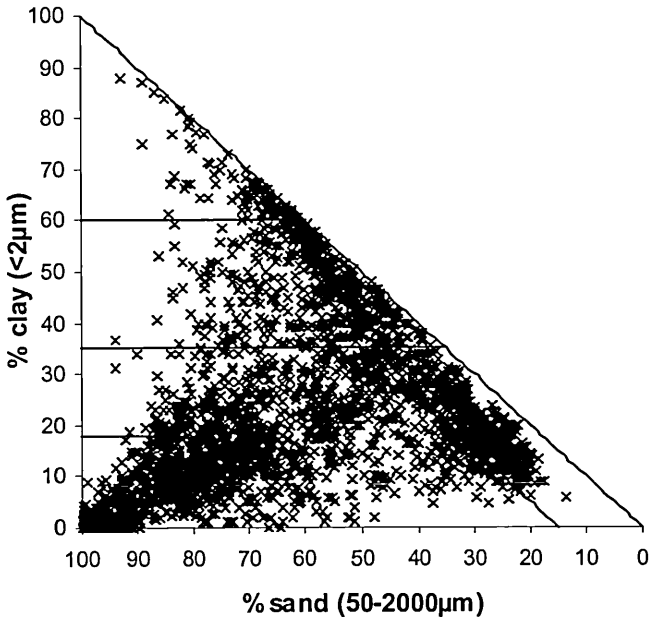



*Figure 5.1d: Class distribution of the data selected according to the selection criteria*

## 5.2 Results of Statistical Analyses of HYPRES

In this section, the HYPRES data that will be used for the modelling process (which will be discussed in some detail in Chapter 6) is analysed using various statistical tests, and the results discussed in general, and specifically where appropriate. Firstly, I will report on results of a test to determine whether the textural components of each soil horizon sum to 100, as expected (see equation 2.2q). The results of such a test will show how noise may be unwittingly added into the database by approximations and erroneous insertion. Further, normality testing will be performed on each of the five input and one output parameters, and detailed by texture class. Following from this, the mean, median, minimum and maximum of each will be calculated, and the standard deviation and coefficient of variation determined. This information will afford an understanding of how the data is distributed, and will prove to be of great benefit when analysing the test results of the ensemble models (Chapter 6).

### 5.2.1 Do the Soil Textural Components Sum to 100?

In terms of the texture component of the soils selected, we should expect that $\Sigma(SSC) = 100$ (see equation 2.2q), but is this the case? Summing the proportions of SSC and rounding to the nearest integer, we find that there are a number of horizons for which this is not the case (Table 5.2a).

| | C | M | MF | F | VF | ALL |
|---|---|---|---|---|---|---|
| Total number of soil horizons (see Table 5.1b) | 630 | 952 | 574 | 516 | 107 | 2779 |
| Number of horizons for which $\Sigma(SSC) = 100$ | 630 | 949 | 567 | 513 | 106 | 2765 |
| Number of horizons for which $\Sigma(SSC) \neq 100$ | 0 | 3 | 7 | 3 | 1 | 14 |
| $\Sigma(SSC)$ when $\Sigma(SSC) \neq 100$ | | 101<br>101<br>102 | 96<br>95<br>99<br>98<br>97<br>99<br>99 | 99<br>101<br>102 | 126 | |

*Table 5.2a: Sum of proportions of SSC, and details where $\Sigma(SSC) \neq 100$*

Of the 2779 soil horizons that were selected for the modelling process, 14 were found to have a mineral content that did not sum to 100. M and F soils had three deviant horizons each, whilst MF and VF had seven and one deviant horizons, respectively. There were no deviant samples in the C texture class of soils. Of all those samples that were deviant, 13 of the 14 were in the range [95, 102]. This may be considered acceptable since these are all within 5% tolerance, however, the deviant VF sample ($\Sigma(SSC) = 126$) certainly cannot be considered acceptable.

Despite the shortcomings of these data, all of these soil horizons have been included in the modelling process. The reason for this is that databases are often found to have errors, some of them large, which contribute to the overall levels of noise (see Section 3.5.3). Although more accurate models may be built with 'clean' data, a model that can also handle errors and noise would be more robust, and a little more realistic.

In addition, in the database many soil horizons can be seen to have SSC integer values. For example, SSC = [40, 40, 20]. Have these values been rounded from, for

example, SSC = [39.6, 40.1, 20.3] or should they be considered to be SSC = [40.000, 40.000, 20.000]?

To the field scientist, approximations are acceptably commonplace, but to the modeller, they represent insoluble noise. One of the stated aims of this thesis is to attempt to inform the field scientist which particular data is of the greatest importance to the modeller and should therefore be measured and handled with care. This is discussed in detail in Chapter 10.

Despite the defects of some data samples in the database, I decided that all horizons should be used in the modelling process, since a certain amount of poor quality data is to be expected in large databases.

### 5.2.2 Anderson-Darling Test for Normality

The Anderson-Darling normality test (e.g. Stephens 1974) performs a hypothesis test to examine whether or not the observations follow a normal distribution.

This normality test provides two goodness-of-fit measures to help assess how well the distribution fits the data set; the Anderson-Darling statistic (designated as $A^2$) and the Pearson correlation coefficient (also termed 'P-Value', or P). $A^2$ is a measure of how far the plot points fall from the fitted line in a probability plot, where smaller values indicate that the distribution fits the data better. In general, if $A^2 > 0.753$, there is 95% confidence that the distribution is non-normal. For P, a coefficient ranging between 0 and 1, greater values indicate a better fitting distribution. In general, if $P > 0.05$, there is 95% confidence that the data set is normally distributed.

Table 5.2b show the results of the Anderson-Darling normality test when applied to the five input (SSC, BD and %OM) and one output ($\theta_{FC}$) parameters of each texture class.

The majority of these distributions are clearly highly non-normal, having values of $A^2$ that are greater than 0.753, and very low values of P, most of which are very much smaller than 0.05. Exceptions are the BD and $\theta_{FC}$ distributions of VF soil, which both have $A^2 < 0.753$ and $P > 0.05$, indicating that they are normally distributed.

| | %Cl | | %Si | | %Sa | |
|---|---|---|---|---|---|---|
| | $A^2$ | P | $A^2$ | P | $A^2$ | P |
| C | 8.0 | 0.000 | 6.8 | 0.000 | 13.7 | 0.000 |
| M | 2.0 | 0.000 | 7.5 | 0.000 | 20.4 | 0.000 |
| MF | 7.4 | 0.000 | 8.5 | 0.000 | 1.9 | 0.000 |
| F | 6.2 | 0.000 | 3.4 | 0.000 | 20.4 | 0.000 |
| VF | 5.8 | 0.000 | 3.0 | 0.000 | 6.6 | 0.000 |

| | BD | | %OM | | $\theta_{FC}$ | |
|---|---|---|---|---|---|---|
| | $A^2$ | P | $A^2$ | P | $A^2$ | P |
| C | 2.1 | 0.000 | 49.7 | 0.000 | 2.0 | 0.000 |
| M | 4.2 | 0.000 | 71.6 | 0.000 | 3.7 | 0.000 |
| MF | 10.2 | 0.000 | 26.6 | 0.000 | 19.8 | 0.000 |
| F | 1.1 | 0.006 | 24.9 | 0.000 | 0.8 | 0.032 |
| VF | 0.7 | 0.083 | 6.8 | 0.000 | 0.6 | 0.111 |

*Table 5.2b: Results of normality testing on the five input and one output parameters chosen for modelling*

### 5.2.3 Minimum and Maximum

Considering that the distributions of most of the individual parameters are highly non-normal, it may prove to be useful to determine their ranges, detailed by texture class, Table 5.2c.

| | %Cl | | %Si | | %Sa | |
|---|---|---|---|---|---|---|
| | Min | Max | Min | Max | Min | Max |
| C | 0 | 17.6 | 0 | 33.5 | 65 | 100 |
| M | 0 | 34.8 | 5.9 | 77.2 | 15 | 72.5 |
| MF | 6 | 34.7 | 29.5 | 86.4 | 2.1 | 43.5 |
| F | 35 | 60 | 6.215 | 62.4 | 0.1 | 56.9 |
| VF | 60.2 | 88 | 7 | 42.2 | 0 | 23 |

| | BD | | %OM | | $\theta_{FC}$ | |
|---|---|---|---|---|---|---|
| | Min | Max | Min | Max | Min | Max |
| C | 0.83 | 1.92 | 0 | 9.8 | 0.01 | 0.44 |
| M | 0.76 | 1.97 | 0 | 9.4 | 0.10 | 0.57 |
| MF | 0.55 | 1.76 | 0 | 9.9 | 0.15 | 0.58 |
| F | 0.60 | 1.93 | 0 | 8.9 | 0.11 | 0.70 |
| VF | 0.54 | 1.57 | 0.1 | 8.6 | 0.37 | 0.76 |

*Table 5.2c: Minimum and maximum values of the categorised parameters*

Comparing these ranges with those defined by the FAO (1990), Table 5.2d (also see Figure 2.1b), may yield important information.

| | %Cl | | %Sa | |
|---|---|---|---|---|
| | > | < | > | < |
| **C** | | 18 | 65 | |
| **M** | 18 | 35 | 15 | |
| **M** | | 18 | 15 | 65 |
| **MF** | | 35 | | 15 |
| **F** | 35 | 60 | | |
| **VF** | 60 | | | |

*Table 5.2d: FAO (1990) criteria for composition of the texture classes*

It should be noted that the definitions are based on values of %Cl and %Sa, and also that the M texture class has two possible definitions.

Comparing Tables 5.2c and 5.2d, the ranges of all texture classes fall within their FAO (1990) defined ranges, with the exception of %Sa for MF soils. For this texture class, the sand content should not be more than 15%; however, the maximum value in the HYPRES MF texture class is %Sa = 43.5. In the database this texture is entered as SSC = [43.5, 29.5, 27]. The total of these proportions sums to 100, so the previous test of $\Sigma$(SSC) = 100 did not discover this deviant soil horizon. According to the textural composition of this particular soil horizon, it is clear that it should not have been designated as an MF soil, but in fact should belong to the M texture class. Nevertheless, this soil horizon is retained as an MF soil, as a further example of how noise may be unwittingly inserted into databases. A more detailed analysis of the %Sa component of MF soils did not yield further horizons for which %Sa > 15.

### 5.2.4 Mean and Median

Now that we know the ranges of each of the input and ouput parameters of each texture class, determining the mean and median of such will allow the investigator to understand where the average values lie, Table 5.2e.

| | %Cl | | %Si | | %Sa | |
|---|---|---|---|---|---|---|
| | **Mean** | **Median** | **Mean** | **Median** | **Mean** | **Median** |
| **C** | 5.94 | 5.75 | 11.69 | 11.51 | 82.37 | 82.49 |
| **M** | 18.82 | 18.75 | 39.09 | 38.47 | 42.09 | 41.99 |
| **MF** | 20.11 | 19.80 | 70.71 | 71.24 | 9.14 | 9.13 |
| **F** | 46.13 | 46.05 | 41.73 | 42.27 | 12.14 | 10.69 |
| **VF** | 66.86 | 65.68 | 29.09 | 29.86 | 4.29 | 3.51 |

| | BD | | %OM | | $\theta_{FC}$ | |
|---|---|---|---|---|---|---|
| | **Mean** | **Median** | **Mean** | **Median** | **Mean** | **Median** |
| **C** | 1.536 | 1.544 | 1.40 | 1.06 | 0.162 | 0.159 |
| **M** | 1.489 | 1.500 | 1.60 | 1.22 | 0.296 | 0.294 |
| **MF** | 1.469 | 1.485 | 1.16 | 1.03 | 0.333 | 0.327 |
| **F** | 1.321 | 1.328 | 1.97 | 1.69 | 0.420 | 0.420 |
| **VF** | 1.177 | 1.183 | 2.15 | 1.79 | 0.496 | 0.493 |

*Table 5.2e: Mean and median of the categorised parameters*

It may be noted that as the %Cl increases (i.e. as you read down the table of texture classes from C to VF), the average $\theta_{FC}$ increases, and also that as the average BD increases, the average $\theta_{FC}$ decreases. This is what we would expect, since as the BD increases, the amount of matter contained in the sample increases, leaving less room for the pore spaces. Thus, since the pore space is where soil water is stored, soils having larger values of BD can retain less water. Additionally, as the %OM increases, $\theta_{FC}$ increases. This is also what we would expect, since in general, soils with greater organic matter content tend to retain more water at the same matric potential than soils with lesser organic matter content. This suggests that $\theta_{FC}$ may be an increasing function of %OM and %Cl, and a decreasing function of BD.

Comparing each of the categorised parameter values for the mean and median (Table 5.2e) with their respective ranges (Table 5.2d) allows an understanding of whether their distributions are right- or left-skewed. To do this, I calculated the distances from the minimum to the maximum values and determined, as integer percentages, where along this line the mean and median lie, Table 5.2f. The mean results have also been plotted, Figure 5.2a (the median has been omitted, since there is no significant difference between the mean and median values for each of the categorised parameters).

| | %Cl | | %Si | | %Sa | |
|---|---|---|---|---|---|---|
| | **Mean** | **Median** | **Mean** | **Median** | **Mean** | **Median** |
| **C** | 34 | 33 | 35 | 34 | 50 | 50 |
| **M** | 54 | 54 | 47 | 46 | 47 | 47 |
| **MF** | 49 | 48 | 72 | 73 | 17 | 17 |
| **F** | 45 | 44 | 63 | 64 | 21 | 19 |
| **VF** | 24 | 20 | 63 | 65 | 19 | 15 |

| | BD | | %OM | | $\theta_{FC}$ | |
|---|---|---|---|---|---|---|
| | **Mean** | **Median** | **Mean** | **Median** | **Mean** | **Median** |
| **C** | 65 | 66 | 14 | 11 | 35 | 35 |
| **M** | 60 | 61 | 17 | 13 | 42 | 41 |
| **MF** | 76 | 77 | 12 | 10 | 43 | 41 |
| **F** | 54 | 55 | 22 | 19 | 53 | 53 |
| **VF** | 62 | 62 | 24 | 20 | 32 | 32 |

*Table 5.2f: Relative position of the mean and median values of parameters, as a percentage of the distance between minimum and maximum*

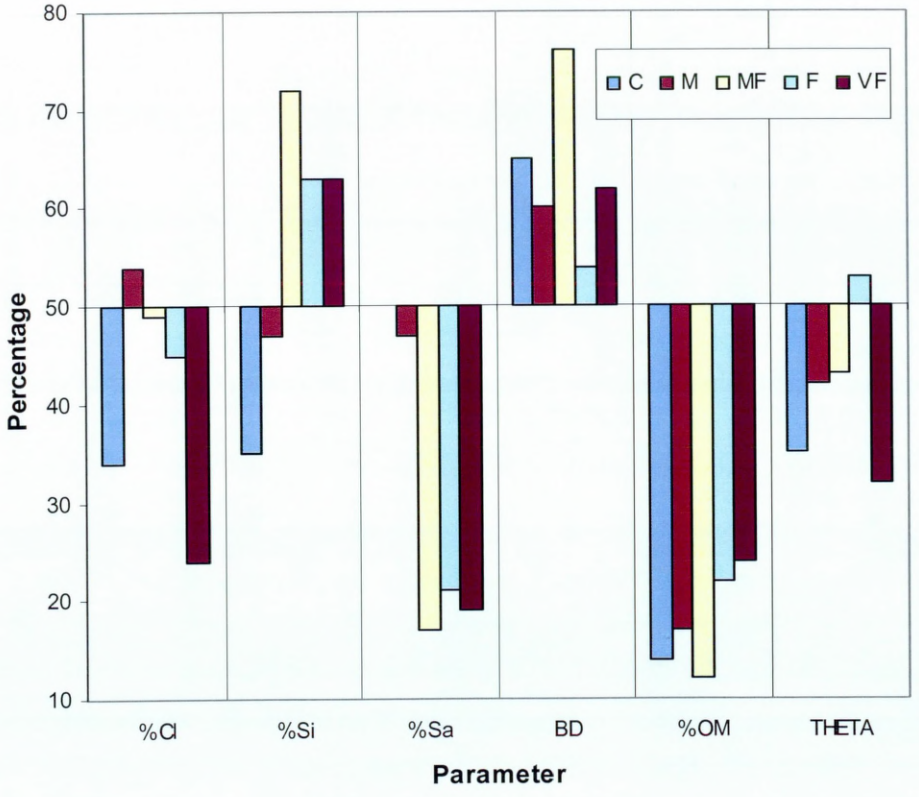



*Figure 5.2a: Mean values of all parameters, plotted as a percentage of the respective distances from their minimum to maximum values*

Low numbers indicate that the mean (or median) lie towards the minimum of the range, whereas higher numbers indicate the reverse. Numbers close to 50 indicate that the average value lies towards the centre of the distribution. Figure 5.2a shows that there is little difference between the mean and median values of the categorised parameters. Additionally, the main features of these graphs are that:

- BD is left-skewed
- $\theta_{FC}$ is mostly right-skewed
- %OM is strongly right-skewed
- %Cl is right-skewed for C and VF soils
- %Si is left-skewed for high proportions of clay
- %Sa is mostly right-skewed, also, as %Cl increases, distribution becomes increasingly right-skewed

### 5.2.5 Standard Deviation and Coefficient of Variation

The standard deviation, $\sigma$, is a measure of the spread of values about the mean, and is expressed as:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} \qquad (5.2a)$$

where $X$ is the $i$th value, for $i = 1$ to $n$, and $\bar{X}$ is the statistical mean.

The coefficient of variation (CV) is a better estimate of the variability, being a measure of the spread of values about the mean, adjusting for the size of the mean, and is expressed as:

$$CV = \frac{\sigma}{\bar{X}} \times 100 \qquad (5.2b)$$

The $\sigma$ and CV of the categorised parameters have been calculated, and are shown in Table 5.2g.

| | %Cl | | %Si | | %Sa | |
|---|---|---|---|---|---|---|
| | $\sigma$ | CV | $\sigma$ | CV | $\sigma$ | CV |
| **C** | 4.54 | 76.5 | 7.93 | 67.8 | 11.00 | 13.4 |
| **M** | 7.74 | 41.2 | 14.77 | 37.8 | 15.89 | 37.8 |
| **MF** | 6.51 | 32.4 | 6.70 | 9.5 | 3.25 | 35.6 |
| **F** | 7.00 | 15.2 | 9.72 | 23.3 | 10.10 | 83.2 |
| **VF** | 6.17 | 9.2 | 7.21 | 24.8 | 4.32 | 100.7 |

| | BD | | %OM | | $\theta_{FC}$ | |
|---|---|---|---|---|---|---|
| | $\sigma$ | CV | $\sigma$ | CV | $\sigma$ | CV |
| **C** | 0.17 | 11.2 | 1.70 | 120.9 | 0.082 | 51.0 |
| **M** | 0.22 | 14.7 | 1.82 | 113.5 | 0.076 | 25.6 |
| **MF** | 0.15 | 10.5 | 1.20 | 103.7 | 0.053 | 15.8 |
| **F** | 0.21 | 15.7 | 1.64 | 83.4 | 0.083 | 19.6 |
| **VF** | 0.19 | 16.3 | 1.79 | 83.3 | 0.066 | 13.2 |

*Table 5.2g: SD and CV of the categorised parameters*

In general, the larger $\sigma$ or CV, the higher is the variability of the values about the mean. In particular, a small value for the $\sigma$ and a high value for the CV indicate that there is a tight spread about the mean, but that those values are highly variable. Additionally, the main features of Table 5.2g are that:

- As %Cl decreases (reading up the table from VF to C) CV of %Cl increases, indicating that, on average, soils having small amounts of %Cl are highly variable in their %Cl contents.
- As %Sa decreases (reading down the table from C to VF) CV of %Sa increases, in particular, the variability of %Sa in F and VF texture classes is very high.
- Variability of %OM is very high, but decreases with increasing %Cl content.
- As %Cl decreases (reading up the table from VF to C), the CV of %Cl and $\theta_{FC}$ both increase. Clay particles, being smaller than silt and sand, tend to fit into the pore spaces created by adjacent particles that are larger than clay. For high clay content soils, many of the larger pore spaces are filled by clay particles, leaving mostly fine, capillary-like pore spaces between the small particles, and water content is high due to the high matric potential. For soils low in clay content, large pore spaces exist, and the sizes and amounts of the clay particles becomes critically important in the filling of these pore spaces. In this case, water content

is low (water drains from the pores readily) and is highly dependent upon the size distribution of clay particles. Thus, for soils having low clay content, the variability in clay content is closely linked to the variability in water content, and therefore samples that have high variability in %Cl are also expected to have high variability in $\theta_{FC}$, and this is exactly what we see.

- As %Sa is increased (reading up the table from VF to C), the CV of %Sa decreases, and the CV of $\theta_{FC}$ increases.

## Chapter Conclusions

Typically, researchers that have constructed ANN-PTFs by the point prediction method have used various combinations of %Sa, %Si, %Cl, BD and %OM as input parameters, and one or more points on the WRC, usually at the inflexion points (field capacity, -33kPa, and/or permanent wilting point, -1500kPa), as the output parameter(s). In this thesis all five parameters (SSC, BD and %OM) are used as inputs to the ANN models. In terms of the output parameter(s), the closest value to the field capacity in HYPRES is the water retention at -250cm$H_2O$ (equal to -24.5kPa). Since the many definitions of field capacity are very vague, the output parameter chosen for modelling is the water retention data held in HYPRES at -250cm$H_2O$, and this is assumed to be the field capacity water retention, denoted $\theta_{FC}$.

Once the relevant input and output parameters have been chosen, and the data extracted from HYPRES, duplicate soil horizons were discarded to reduce noise, and, since there are few organic soils in the database, only mineral soils are considered (those that have %OM < 10) in order to reduce model bias.

Of those soil horizons chosen for the modelling process, statistical tests applied to the five input (SSC, BD and %OM) and one output ($\theta_{FC}$) parameters of each texture class showed that the majority of these distributions are highly non-normal and highly variable.

# Chapter 6

# Continuous PTF of Soil Water Retention – an ANN Ensemble Method

Previous chapters have discussed the necessary background that is required in order to construct a continuous ANN-PTF of soil water retention at field capacity by the ensemble method. This chapter will show how that knowledge may be used to determine optimum parameters for such a model, both independently and, where necessary, dependently. Once the optimum parameters have been established, two new terms will be introduced with which to describe results of tests conducted on the ensemble, and a fully optimised ensemble will be built and tested. Sensitivity analyses will be conducted on the ensemble to determine which of the input parameters are of most importance to such a model. Additionally, an ANN-PTF by the single ANN approach will be constructed, which will provide a comparison between the two methods. These comparisons will show that the ensemble method requires typically one-third or less the amount of data to produce results that are equal to those of the single ANN method. Finally, I will discuss in detail two possible methods of obtaining errors on individual predictions, both theoretically and practically, and will discuss the results of the two competing methods in conjunction with the results of statistical analyses conducted in Section 5.2.

## 6.1 Preliminary Investigations and Optimisation Procedures

There are many parameters that may be altered within the modelling process, many of which, when handled poorly, may lead to the production of erroneous results. It is of prime importance that these parameters are considered carefully in order to optimise the whole modelling procedure. These are discussed in some detail here, from diligent data selection, to a meticulous ANN modelling procedure, and thorough testing.

### 6.1.1 Data Selection

In any artificial intelligence learning system, the accuracy (or competence) of the system will improve with the amount of training data available, and will follow a

learning curve such as that shown in Figure 6.1a. Up to a certain point, additional training data will produce appreciable increases in accuracy, but beyond, additional data will produce little, if any, increase in accuracy. This point in the graph is termed the 'knee point', and represents the point at which the learning system has seen a useful cross-section of data samples that provide adequate coverage of the whole data space or problem domain. The amount of data required for adequate coverage may be very difficult to determine *a priori*, and reflects the complexity of the input-output relationship being modelled, the predictiveness of the input features and the amount of noise in the data (Cunningham *et al.* 2000). Determining the knee point is clearly a useful criterion that would allow the optimum amount of training data to be used, and also ensures that a sufficiently complex dataset remains for the testing phase of the modelling process.

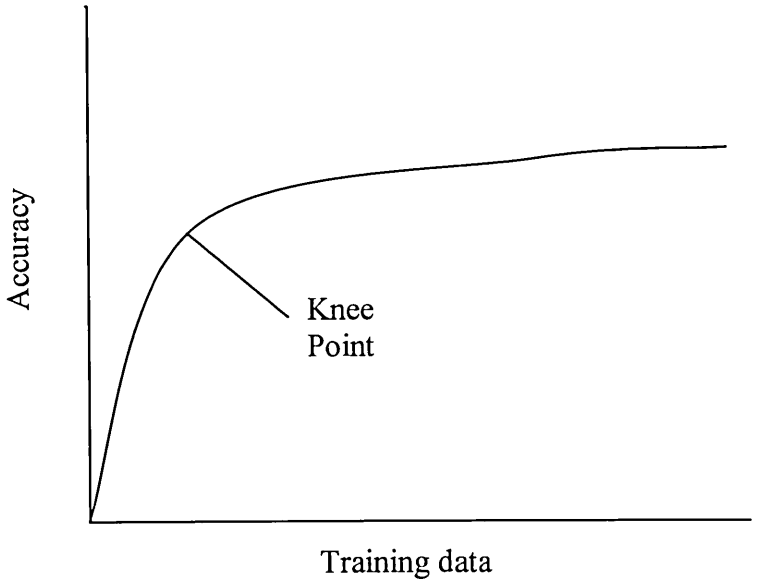



*Figure 6.1a: Typical learning curve plotting accuracy against data*

Choosing ANN training data randomly often leads to high bias in the model (bias/variance trade-off, see Section 4.2.1). Thus, some way to choose an appropriate spread of data must be found (low bias, high variance). The HYPRES database has natural divisions incorporated into it; each soil horizon has been labelled to belong to one of five mineral texture classes (see Figure 2.1a). In the construction of a continuous PTF, if training data were to be chosen randomly from the 2779 horizons extracted from HYPRES, we would expect that there would typically be nine times as

many M horizons as VF horizons (see Table 5.1b). This would lead to a clear bias in the model towards M soils (as compared with VF soil horizons). This is undesirable, since we seek to minimise bias in the model, so initially an equal number of horizons should be chosen from each of these categories to reduce bias towards any particular texture class. Determining exactly how many soil horizons will take us to the knee point is discussed in Section 6.1.7.

### 6.1.2 Re-scaling Algorithm

It is standard practice to perform some pre-processing on any data prior to introducing it to an ANN. This is because of the nature of the transfer function of the neurons. Considering the logarithmic sigmoid (logsig) transfer function as an example (Figure 3.3a), a large input (either positive or negative) would place the output in the saturated region of the transfer function, meaning that a large change of input is required to effect a small change of output. This would mean a correspondingly large amount of time required for the training process, which is clearly undesirable. Thus, for the logsig transfer function, it is necessary to re-scale the data to fall within the range [-1, 1] to avoid saturation. In fact, many ANN practitioners prefer to re-scale their data to be within a range that is *comfortably* within the more linear portion of the plot (i.e. not saturated).

There are an infinite number of possible ways of re-scaling the data, and it is worth remembering that, given the data, some re-scaling algorithms are more suitable than others. I present two possible re-scaling algorithms as examples of how and how not to re-scale the data.

**A logarithmic re-scaling algorithm**

It is quite common for datasets to be re-scaled by taking the logarithm to the base 10 (denoted $\log_{10}$), since this has the effect of 'squashing' the data into a tight range around unity for values up to 100. The only problem here is that $\log_{10}0$ is not mathematically defined, and there are many soil horizons that have a textural make-up of, for instance, SSC = [60, 40, 0], or have %OM = 0. Nevertheless, we could proceed on the basis that the number 0.01 is small in comparison with all the non-zero values for SSC or %OM. It is therefore practical to replace all zeros with the value 0.01 prior

to taking the $\log_{10}$ of the entire dataset. To test whether this is feasible, an ensemble of 21 ANNs was trained to the same error goal. Each ANN had as their input matrix the five parameters SSC, BD and %OM, and 25 soil horizons (five from each of the five mineral texture classes) passed to it as part of the training regime. All zeros in the input matrix were replaced by 0.01, and the matrix re-scaled by taking the $\log_{10}$. The output was the water retention at 0cm$H_2O$ and, since all water retention values are in the range [0, 1], no pre-processing of the output is required. The training data passed to each ANN was different and independent from the data passed to the other ANNs (leading to low bias, high variance). All other ANN parameters, such as ANN architecture, training algorithm, etc., were kept the same for all individual ANNs. When training was complete, each of the individual ANNs were tested by passing to them the values of the input parameters of a soil horizon that they have not previously encountered. The individual ANNs were each required to make a prediction as to what the output (water retention) should be, based on the knowledge they gained from training. These individual predictions were then averaged to give a single prediction, and the process repeated over many soil horizons. The model predictions of water retention were compared to the expected values of water retention for those horizons to give a measure of the error.

The test results are shown in Table 6.1a, broken down into soil texture classes, and overall.

| | C | M | MF | F | VF | ALL |
|---|---|---|---|---|---|---|
| **Number of tests** | 395 | 615 | 376 | 370 | 70 | 1826 |
| **RMSE** | 0.050 | 0.031 | 0.018 | 0.034 | 0.028 | 0.034 |

*Table 6.1a: Test results of an ANN ensemble by texture class and overall*

Comparing the test results by texture class, it can be seen that the errors for the C texture class are significantly higher than those for all other texture classes. What sets C soils apart from soils belonging to the other texture classes is that C soils, by definition, have low proportions of clay, and in many cases, a zero percentage of clay content. At the pre-processing stage, these zeros were altered to 0.01, and it appears that this may have an effect on the RMSE for the C texture class. On closer inspection, the results of individual tests suggest that this is indeed the case. A subset

of selected individual test results of C soil horizons may be seen in Table 6.1b to illustrate this.

| Input parameters | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **%Cl** | **%Si** | **%Sa** | **BD** | **%OM** | $\theta$ | $\hat{\theta}$ | **Hi** | **Lo** |
| 10.4 | 13.98 | 75.62 | 1.192 | 4.49 | 0.54 | 0.52 | 0.63 | 0.30 |
| 8.1 | 11.52 | 80.38 | 1.37 | 4.09 | 0.47 | 0.47 | 0.63 | 0.30 |
| 5.8 | 23.78 | 70.42 | 1.553 | 1.09 | 0.41 | 0.42 | 0.52 | 0.31 |
| 0.01 | 1.5 | 98.5 | 1.667 | 0.17 | 0.37 | 0.24 | 0.70 | -1.12 |
| 0.01 | 1.1 | 98.0 | 1.685 | 0.59 | 0.36 | 0.30 | 2.50 | -1.60 |

*Table 6.1b: A selection of test results (C soils) for an ANN ensemble where $\theta$ and $\hat{\theta}$ represent expected and ensemble-predicted values of the soil water retention, respectively. Hi and Lo represent the highest and lowest single ANN predictions*

The first three rows of results, where the soil sample has a non-zero clay component, show that there is good agreement between the ensemble predictions and expected values for the soil moisture retention, with a reasonable range of single ANN predictions (i.e. all predictions fall in the range [0, 1]). Where the soil sample does not have a clay component, the final two rows of results, there is much poorer agreement. Moreover, the range of values for the predictions shows some poor predictions – recall that $\theta$ should be in the range [0, 1].

Considering the C test results overall (Figure 6.1b) it can be seen that there are many deviations from the 1:1 correlation (dotted line) that one would hope to see between the expectation and the ensemble prediction of the water retention. In addition, there is a considerable difference between the 1:1 correlation that one would expect and the regression line (solid line) shown.

This is typical of test results for this model – clearly, replacing a zero with a 0.01 and re-scaling with a logarithm at the pre-processing stage is not practical.

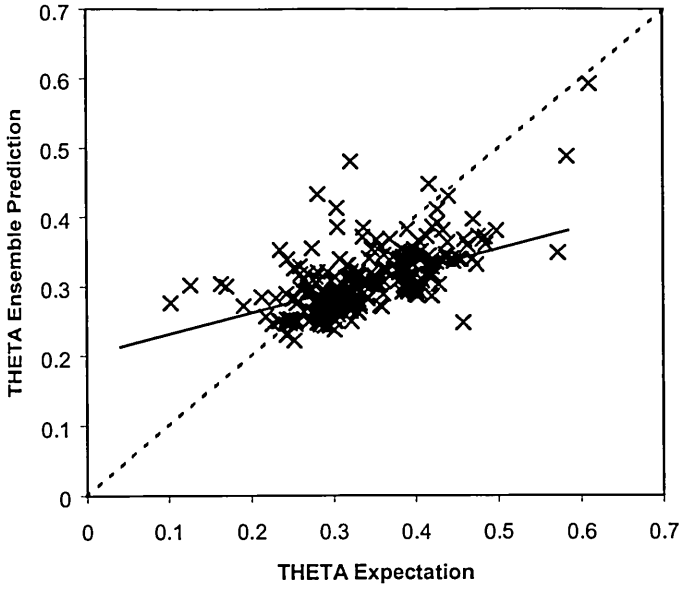



*Figure 6.1b: Correlation between expected and ensemble-predicted values of $\theta$*

**A cosine re-scaling algorithm**

What is required then, is an algorithm that will re-scale numbers from [0, 100] (i.e. the effective range of the values of all input parameters) into the range [-1, 1] and will handle the number zero as effectively as any other value.

The cosine function shown graphically in Figure 6.1c has a unique output for an input range of [0, $\pi$] (measured in radians) and does not discriminate against inputs of zero. It would seem, therefore, to be a good alternative to the $\log_{10}$ algorithm.

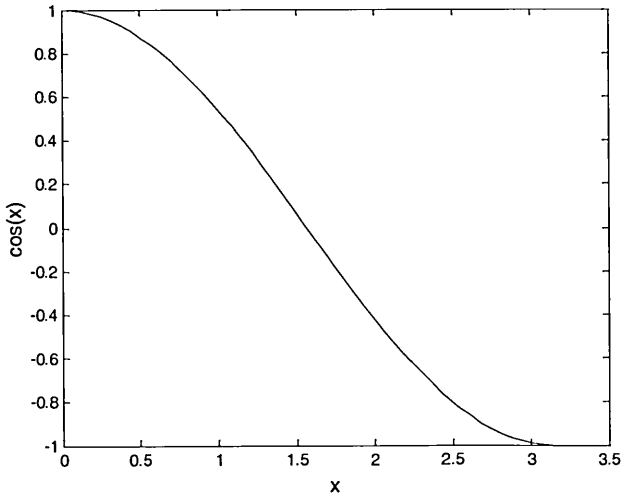



*Figure 6.1c: Cosine curve for the input range [0, $\pi$]*

Since the input of the data is in the range [0, π], it is necessary to multiply the input matrix by π/100 to utilise the entire range of the first half of the cosine curve. Also, since inputs of 0 and π would yield outputs of 1 and –1, respectively, the inputs should also be multiplied by –1, so that the same inputs now yield the more appropriate outputs of –1 and 1.

In summary, the input matrix should be re-scaled by the algorithm:

$$u_j = \cos\left\{-\frac{\pi}{100}\left(input_j\right)\right\} \tag{6.1a}$$

To determine whether this algorithm is preferable to the $\log_{10}$ algorithm, four datasets were constructed. These comprise of two sets of data for use in the training phase, termed TRAIN_1 and TRAIN_2, and two sets of data for use in the testing phase, termed TEST_1 and TEST_2. Both training sets comprise five soil horizons from each soil texture class. TRAIN_1 and TEST_1 are datasets that contain no zeros, whereas TRAIN_2 has %Cl = 0 for all five C soil horizons and TEST_2 also contains some C textures for which %Cl = 0, and some horizons for which %OM = 0.

Both of the training datasets were duplicated and modelled by a single ANN, one with the $\log_{10}$ re-scaling algorithm and the other with the cosine algorithm of Equation 6.1a. All zeros were altered to 0.01 in the case of the $\log_{10}$ re-scaling algorithm, but unchanged for the cosine algorithm, and all other training parameters, such as ANN architecture, training algorithm, etc., were kept the same for the modelling process. This resulted in four ANNs, each of which were tested with TEST_1 and TEST_2. In addition, each of the ANNs were tested with seen data.

The results of testing are shown in Table 6.1c.

| | | | RMSE | |
|---|---|---|---|---|
| | **Algorithm** | **Accuracy** | **TEST_1** (no zeros) | **TEST_2** (zeros) |
| **TRAIN_1** (no zeros) | Cos | 0.008 | 0.041 | 0.037 |
| | $\log_{10}$ | 0.007 | 0.129 | 0.084 |
| **TRAIN_2** (five zeros) | Cos | 0.006 | 0.053 | 0.075 |
| | $\log_{10}$ | 0.007 | 0.064 | 0.052 |

*Table 6.1c: Test results illustrating the difference between the logarithm and cosine algorithms on re-scaling the input data*

There are a number of things to note from these results:

- The accuracy of all four ANNs is approximately the same.
- TRAIN_2 ($\log_{10}$) < TRAIN_1 ($\log_{10}$) for TEST_1 and TEST_2.
  - This suggests that ANNs trained with a dataset that has zeros (this is the minority case) is more reliable than ANNs trained with a dataset that has no zeros, using the $\log_{10}$ algorithm, regardless of whether the test set has zeros.
- TRAIN_1 (cos) < TRAIN_2 (cos) for TEST_1 and TEST_2.
  - This suggests that ANNs trained with a dataset that has no zeros (this is the majority case) is more reliable than ANNs trained with a dataset that has zeros, using the cos algorithm, regardless of whether the test set has zeros.
- TRAIN_1 (cos) < TRAIN_1 ($\log_{10}$) for TEST_1 and TEST_2.
  - This suggests that the cos algorithm is preferable to the $\log_{10}$ algorithm for ANNs with no zeros in the training set (this is the majority case), regardless of whether the test set has zeros.
- TRAIN_2 (cos) < TRAIN_2 ($\log_{10}$) for TEST_1.
  - This suggests that the cos algorithm is preferable to the $\log_{10}$ algorithm for an ANN with zeros in the training set, and no zeros in the test set.
- TEST_1 ≈ TEST_2 for TRAIN_1 (cos).
  - For a training set containing no zeros, the cos algorithm is equally reliable when there are zeros in the test set as not.

The conclusion here is that, for the test cases presented, the cosine algorithm out-performs the $\log_{10}$ algorithm in almost all cases. As such, all subsequent investigations utilise this cosine algorithm, albeit in a slightly modified form:

$$u_j = \cos\left\{-\frac{\pi}{100}\left(0.8 \times input_j + 0.1\right)\right\} \tag{6.1b}$$

This modification helps to keep the extremes of the inputs to within 10% of saturation at both ends of the transfer function. The error surface (see Section 3.5.2) associated with a given dataset is fixed for that dataset, and cannot be altered (unless the dataset is changed, added to or reduced); it can only be traversed, hopefully

towards a global minimum. Thus, this modification does not impact on the final output or the error function, but rather allows the ANN to reach optimisation quicker than if the transfer function were to be operating in a region of saturation.

### 6.1.3 Optimum Number of Test Sets

On testing ANNs or ensembles, it is useful to know how many tests are required to be performed to give an indication of how much confidence you have in the results. When testing the ANN ensemble trained with the $\log_{10}$ algorithm, the running RMSE of the results for each texture class was also calculated.

To determine a running RMSE, perform test #1, determine the RMSE and plot the RMSE against the number of tests performed. Perform test #2, determine the RMSE of tests 1&2, and plot. Repeat this procedure for all tests.

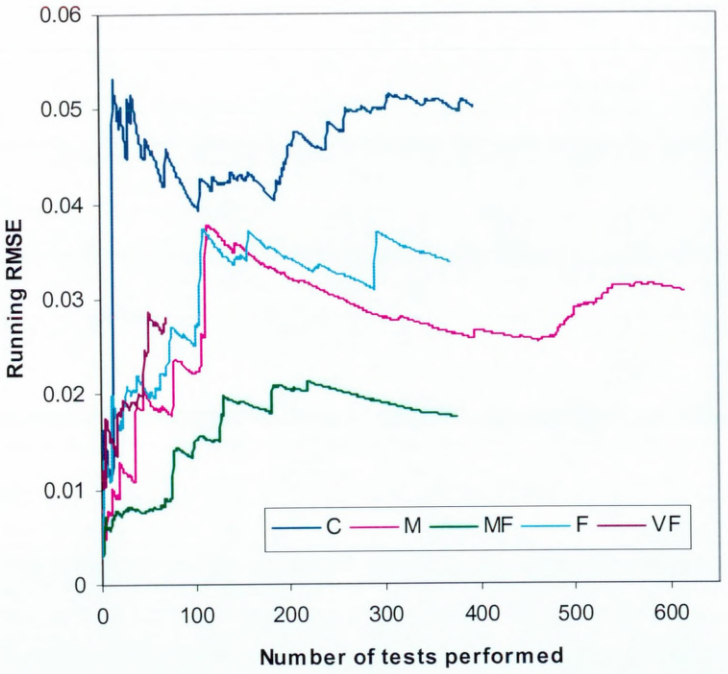



*Figure 6.1d: Graphs of running RMSE versus number of tests performed*

Graphs of running RMSE *versus* number of tests performed (Figure 6.1d) show that the RMSEs appear to approximate a steady-state[1] condition at the two decimal point level after approximately 150-200 tests for all texture classes with the exception of VF soils.

Since there were only 70 tests in the VF soil horizons texture class, it would be expected that the graph of running RMSE *versus* number of tests performed for this texture class would not reach steady-state at the two decimal point level, and this is indeed the case.

Thus, in order to ensure that testing has been stringent, it is necessary to perform of the order of 150-200 tests (or preferably more) per texture class in the case of continuous PTFs when one wishes to report RMSEs that are accurate to two decimal places.

### 6.1.4 Optimum Number of Ensemble Members

In an earlier test, 21 ANNs were used in an ensemble, but is this the optimum number of members to use?

To find the optimum number of ANNs to use in an ensemble, each ANN was re-trained with the cosine algorithm, and each of the 21 ANNs tested independently. Each independent prediction was recorded. This yielded 21 sets of predictions, each corresponding to a given ANN. The weighted average of two sets of predictions to effect the predictions of a two-membered ANN ensemble was calculated, and similarly with three sets and four sets, etc., up to and including 21 sets. In this way ensembles with increasing numbers of members was produced. In addition, this procedure was repeated five times with randomly chosen members from the population to facilitate smoothing of the RMSE test results.

The RMSE of a given ensemble was then plotted against its number of ANN members, and the results of this investigation are shown in Figure 6.1e.

[1] Steady-state has many different definitions, depending on the scientific discipline to which it is applied. In the context of this thesis, steady-state is defined to be a stable condition that exhibits only negligible change within a given level of tolerance over an arbitrarily long period of time or an arbitrarily large number of tests.

A statistical measure, termed the 'relative improvement' (RI) can be used to assess the relative improvement of one result over another, and may be defined in terms of the RMSE as:

$$RI = \frac{RMSE(x) - RMSE(y)}{RMSE(x)} \times 100 \qquad (6.1c)$$

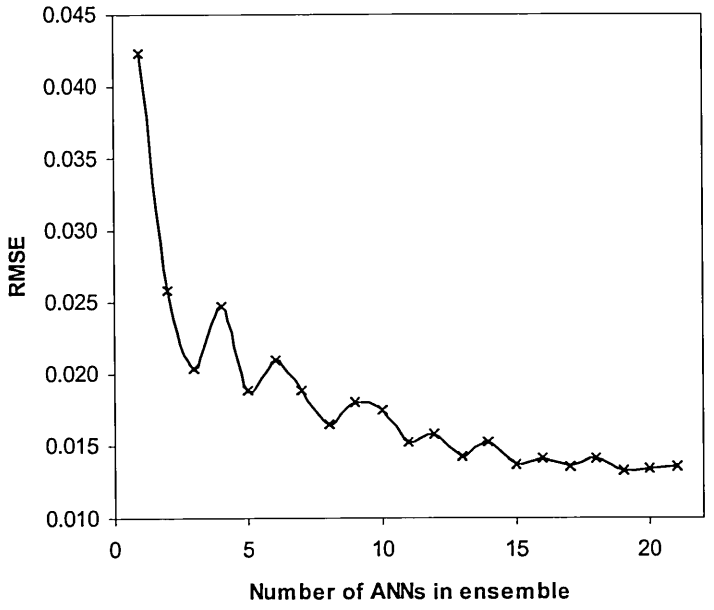



*Figure 6.1e: Graph of RMSE* versus *number of ANN ensemble members*

Calculating from Figure 6.1e, the relative improvement of 11 ANN ensemble members over one member is 65%. The relative improvement of 21 members over one member is 67%. Clearly, the extra work involved in constructing more ANN members over the first 10 or 11 is not worth the tiny benefits on offer.

It is also worth noting that the relative improvement of two ANN members over one member is almost 40%, suggesting that ensembles offer greatly improved results over single ANN methods, even when the number of members is as few as two.

### 6.1.5 Training Parameters and Algorithms

In Section 3.5.2 I discussed in some detail a selection of appropriate training algorithms that may be used with regression-based problems such as the one faced

here. The overall sentiments regarding training algorithms are that *when used correctly* there should not be significant differences in models constructed with different training algorithms. After all, the error surface (see Section 3.5.2) associated with a given dataset is fixed for that dataset, and cannot be altered (unless the dataset is changed, added to or reduced); it can only be traversed, hopefully towards a global minimum. Different training algorithms allow different ways of exploring the error surface, some more quickly, and some more efficiently than others. The aim of the investigator, therefore, is to find the algorithm that is most appropriate for his/her needs. Here I placed a stress on the phrase 'when used correctly'. This is important, since there are issues to be resolved, such as underfitting and overfitting (see Section 3.6.4), setting appropriate error goals and learning rates (see Section 3.5.2).

In Sections 6.1.2 to 6.1.4 the results of investigations aimed at finding optimal levels and appropriate algorithms of various parameters to be used in the final PTF model were discussed. All of these investigations used the BP algorithm, and since all other parameters were held constant (i.e. size of dataset, error goal, learning rate, etc.), it was of no consequence that the training may have under/overfitted the data, or that a minimum, whether local or global, may not have been reached. The investigations were carefully designed so that only the specific parameter or variable under examination was varied or allowed to vary, and thus, optimal values may be found. In Section 3.5.2 it was observed that the BR algorithm would provide the quickest means of reaching a global minimum, whilst generalising more accurately than any other of the discussed algorithms. This is the algorithm that will be employed in all subsequent investigations in this thesis.

### 6.1.6 Evidence for the Optimum ANN Architecture

In Section 3.5.2 I discussed in detail how the BR algorithm may assist in choosing the most appropriate ANN architecture – the main advantage being that it is a model-driven, not a data-driven algorithm. Investigations in Sections 6.1.2 to 6.1.4 were all conducted with a 6-4-1 architecture (6 neurons in the input layer, 4 in the hidden layer and a single output neuron) after Minasny and McBratney (2002). However, is this the optimum architecture?

Firstly, it is prudent to consider the dataset that we are investigating. It is usual for the number of input neurons to be matched to the number of input parameters, and the

number of output neurons matched to the number of output categories required. Since we wish to model SSC, BD and %OM at the input stage, and $\theta_{FC}$ at the output stage, it is clear that we should construct a 5-$n$-1 ANN, where $n$ represents the number of neurons in the hidden layer.

At the training stage, the BR algorithm gives an output that indicates how many of the ANN parameters (weights and biases) are being effectively used by the ANN. Since ANNs containing too many neurons will be pruned (see Section 3.5.2) to the optimal architecture, the number of ANN parameters being used gives an indication as to how many hidden neurons should be chosen. Since the most appropriate architecture of the ANN used here is 5-$n$-1, it is useful to calculate how many possible ANN parameters there are in the ANN for various numbers of hidden neurons. Here, $W$ and $b$ represent ANN weights and biases, respectively, and subscripts $i$, $h$ and $o$ represents the input, hidden and output neurons, respectively.

- **Input ANN parameters.** There are five input parameters connected to each of five input neurons, giving 25 weights, plus, each of the five neurons has an associated bias, therefore:
  - $W_i + b_i = 25 + 5 = 30$

- **Hidden ANN parameters.** There are five inputs neurons connected to each of $n$ hidden neurons, giving $5n$ weights, plus, each of the $n$ neurons has a bias, therefore:
  - $W_{ih} + b_h = 5n + n = 6n$

- **Output ANN parameters.** There are $n$ hidden neurons connected to one output neuron, giving $n$ weights, plus, the output neuron has a bias associated with it, therefore:
  - $W_{ho} + b_o = n + 1$

The total number of ANN parameters ($P_{init}$) is the sum of the parameters for each layer, and is thus:

$$P_{init} = 31 + 7n \qquad (6.1d)$$

Table 6.1d details how many ANN parameters are present in a 5-*n*-1 architecture ANN, for different numbers of neurons in the hidden layer.

| *n* | **Architecture** | $\mathbf{P_{init}}$ |
|---|---|---|
| 1 | 5-1-1 | 38 |
| 2 | 5-2-1 | 45 |
| 3 | 5-3-1 | 52 |
| 4 | 5-4-1 | 59 |
| 5 | 5-5-1 | 66 |

*Table 6.1d: Number of initial ANN parameters in a 5-n-1 ANN*

During the training phase, the BR algorithm gives an output to the user, showing the effective numbers of ANN parameters that are being used, $P_{train}$. The ANN output $P_{train}$ can be directly compared with $P_{init}$ to determine the optimal ANN architecture. For example, if on a trial training run we chose a 5-3-1 ANN architecture, this would give $P_{init} = 52$. At the end of the training period, if $P_{train}$ is, say, 32.3, it is clear that, although the ANN would never overfit the data, it would take longer to train an ANN that is overly complex for the problem than one that is sufficiently complex, and we should alter the architecture from a 5-3-1 to a 5-1-1.

It is clear that as the amount of training data passed to a given ANN is increased, the complexity of the dataset as a whole is increased. This leads in turn to an increase in the complexity of the function required to model the data, and a concomitant increase in the number of hidden neurons required to sufficiently describe the level of complexity. Thus, the optimum number of hidden neurons to use depends on the number of soil horizons passed to the ANN on training, and these should be determined together empirically.

### 6.1.7 Evidence for the Optimum Number of Soil Horizons per ANN

To determine the optimum number of hidden neurons and the optimum number of soil horizons to use per ANN, nine test ANNs, each with an architecture of 5-10-1 and numbered from one to nine were constructed. ANN_1 had 25 soil horizons passed to it for training, five horizons from each of the five soil texture classes. Subsequent ANNs had a further 25 soil horizons passed to it, so that ANN_2 trained with 50 horizons, ANN_3 trained with 75 soil horizons, and so on, Table 6.1e .

$P_{train}$ for each ANN ranged from 11.6 for ANN_1 to 25.5 for ANN_9. Now since $P_{init}$(5-1-1) = 38, it is clear that $P_{init}$(5-1-1) > $P_{train}$(for $n$ = 1 to 9), and thus the optimum architecture for ANNs containing up to 225 soil horizons (at $\theta_{FC}$) is 5-1-1.

| ANN number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Number of soil horizons per texture class | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 |
| Total number of soil horizons | 25 | 50 | 75 | 100 | 125 | 150 | 175 | 200 | 225 |

*Table 6.1e: Number of soil horizons passed to the nine test ANNs*

On setting each ANN to be of the architectural form 5-1-1, training was conducted on each of the ANNs until the BR algorithm determined that the effective number of parameters had converged. This is a good indication that the algorithm has truly converged and the optimisation process is complete, at which stage the training stops automatically. Testing of each of the nine ANNs was conducted for both accuracy (25 test sets) and reliability (100 test sets), and the results are shown in Figure 6.1f. Recall that when a trained ANN is tested with data that it has previously encountered, then the data is referred to as 'seen', and the results of testing are called 'accuracy'. Conversely, when a trained ANN is tested with data that it has not previously encountered, then the data is referred to as 'unseen', and the results of testing are called 'reliability'.

Firstly, it should be noted that the results of tests to determine the accuracy are always better than those for reliability. This is to be expected, since the accuracy is a reflection of the ANNs' ability to produce predictions for which they *already have the answer*. In other words, if a field scientist wished to know the soil hydraulic properties of a new soil sample, one that has a duplicate in the database whose properties have already been measured, it would be quicker and easier to simply look it up in the database rather than request a prediction from a model. This is specifically why we are not interested in the accuracy results.

In terms of the reliability plot, what we are interested in is the general trend of the results to provide evidence as to the optimum number of horizons to use per ANN ensemble member. The investigation to produce the plot could have been reproduced

a number of times to facilitate smoothing (very time consuming), or alternatively, a best fit trendline could be added to the plot to show the general trend.

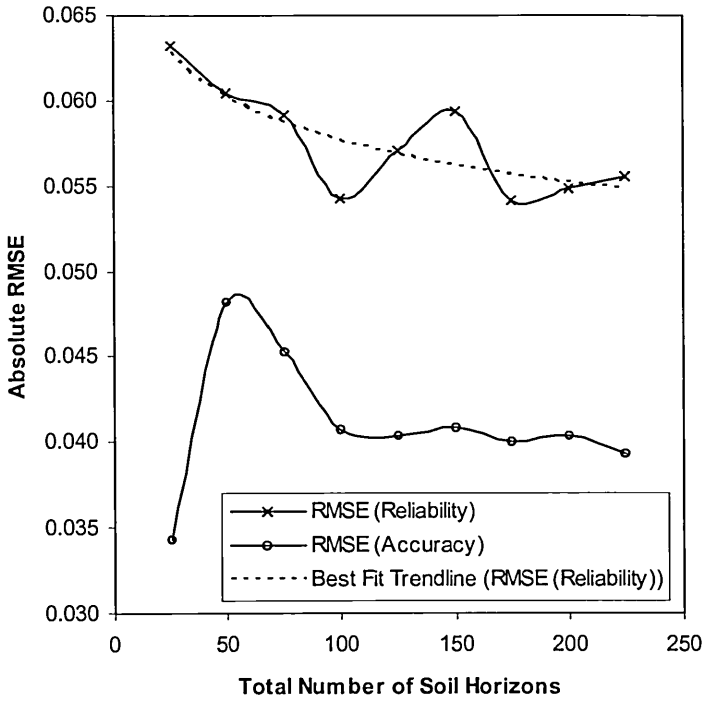



*Figure 6.1f: Results of tests to obtain evidence for the optimum number of soil horizons per ANN*

The reliability best fit trendline continues to decrease from ANN_1 to ANN_9, but the rate of decrease slows down as more soil horizons are added. This suggests that the more information that the ANN has, the more closely it may approach the underlying function of the data. This is what we would expect. There is a limit, however, to the amount of additional information that can be passed to the ANN; once it has all the information required to represent the function, any further information will not result in an increased predictive ability. For this reason, it is to be expected that the general trend of the reliability results will tend towards steady-state as more soil horizons are added to the model.

Considering the best fit trendline fitted for the reliability test set, the results seem to suggest that significant improvement may be made over an ANN with 25 soil horizons (ANN_1), but that only small improvements may be made with greater than about 125 horizons (ANN_5). Quantitatively, the relative improvement of 125

horizons over 25 horizons is 9.8%, whereas the relative improvement of 225 horizons over 25 horizons is 12.7%. Contrast this latter figure with the relative improvement of two ANN ensemble members over one from tests conducted earlier – almost 40%. The small improvement on (almost) doubling the number of soil horizons (from 125 to 225) passed to a single ANN pales when compared to the improvements that could be made by passing half of the data to each of two ANN members of an ensemble. This point will be re-visited in Section 6.2.2.

It is clear then, that the additional work (not to mention the additional constraints on the amount of data required) of constructing ANNs with more than about 125 horizons is not really worth doing, when an ANN trained with 125 soil horizons produces a more than satisfactory result, compared to an ANN with only 25 horizons.

## 6.2 Continuous PTF – Results and Discussion

Section 6.1 detailed the various parameters that need to be optimised in order to build an ANN-PTF by the ensemble method that generalises well for unseen data, and maximises information usage, whilst minimising the amount of data that is required. In this section I will show that it is necessary to introduce two new terms with which to describe PTFs that are constructed using the ensemble method. I will also construct a number of ANN-PTF ensembles, utilising all the lessons that have been learnt in Section 6.1, and perform sensitivity analyses to determine which input parameters are of most importance to the model.

### 6.2.1 Green and Durable – New Terms for ANN-PTF Ensembles

In Section 2.5 I introduced the commonly used terms of 'seen' and 'unseen' when referring to testing ANNs. However, this terminology has only previously been used for describing data in the single ANN method, and is not sufficient when discussing the ensemble method. To qualify this, consider an example of five ANNs, termed *A*, *B*, *C*, *D* and *E*, and each of these ANNs is trained using five data patterns chosen randomly from 26, designated from *a-z*. One possible combination of training patterns for these five ANNs is given below, and also consider all ANNs to be tested by, say, *a*, *b* and *c*, and the predictions combined to produce an ensemble prediction.

| | ANN | | | | |
|---|---|---|---|---|---|
| | ***A*** | ***B*** | ***C*** | ***D*** | ***E*** |
| **Training patterns** | *a t k c o* | *w l c p q* | *j u t n c* | *g t l c d* | *m x f c s* |
| **Test patterns** | *a b c* | *a b c* | *a b c* | *a b c* | *a b c* |
| **Seen data** | *a c* | *c* | *c* | *c* | *c* |
| **Unseen data** | *b* | *a b* | *a b* | *a b* | *a b* |

Test *c* would be considered as seen by all ANNs, test *b* would be unseen by all ANNs, and test *a* would be seen by ANN *A*, but unseen by all other ANNs. In terms of the ensemble, test *c* is seen, and its results could be described as the accuracy. Test *b*, is unseen by the ensemble, and is thus described as the reliability. However, test *a* is both seen and unseen by the ensemble, and as such, cannot correctly be described as either, and neither can its results be regarded as the accuracy or reliability. In this case, though, the result would more closely resemble the reliability, since the data is unseen by four-fifths of the ANNs in the ensemble.

It is proposed then that, for ensembles, tests with data that are both seen and unseen, i.e. used in both training (of at least one member ANN) and test datasets, should be described as 'green', with reference to the fact that the data has been recycled (used more than once). Further, results of tests with green data will henceforth be referred to as the 'durability' of the test procedure.

I established in Section 6.1.7 that we are not interested in accuracy results, as they are results for which we already have the answer. For single ANNs, the reliability is what we are interested in, and for ensembles, when the number of ANN members is large, since the durability of the ensemble is closely related to the reliability, the durability is what we will now be concerned with.

Many ANN-PTF practitioners have in the past insisted on maintaining complete independence between data for training and data for testing to preserve the integrity of the reliability results. For a single ANN method, this is good practice and should be encouraged. However, in the ensemble method, provided that the amount of data for testing is significantly larger than that for training and the number of member ANNs

is large, this is not necessary, since most tests will be unseen, and the results of green tests will very closely resemble those of unseen. As an example of this, consider ten ANNs trained to optimisation by using 25 soil horizons per texture class sampled randomly with replacement from the available data. These ANNs are then combined to create an ensemble, of one, two, ..., ten members. The resulting ensembles are tested with all available data, and results detailed by durability and overall, as well as by texture class. A plot of RMSE *versus* number of ensemble members for C soils is shown in Figure 6.2a.

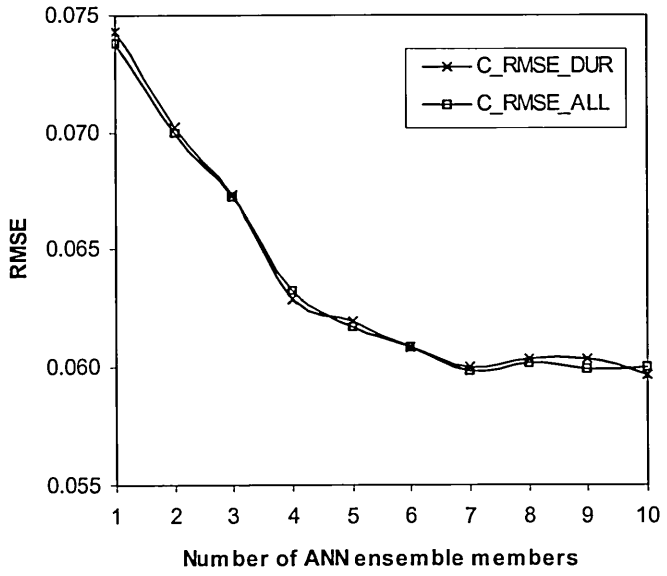



*Figure 6.2a: RMSE* versus *number of ensemble members for C soils*

The durability (C_RMSE_DUR) of each of the ensembles very closely resembles the results of all tests (C_RMSE_ALL). This is not surprising, since the green tests are virtually independent from the training data. To understand this, consider that there are 630 C soil horizons in the dataset that may be selected for training. Since this dataset is Bootstrapped 25 times to make up a dataset for training, and there are ten ANN ensemble members, the probability that an individual ANN member in the ensemble encounters a given C sample is 25 x ($^1/_{630}$) x ($^1/_{10}$), i.e. less than 0.004. This suggests that the probability that any given C soil horizon occurs both in training and testing of the ensemble is less than 0.4%. In other words, for any given C soil horizon, the training and test datasets are more than 99.6% independent on average. Even with

the VF texture class, where there are only 107 soil horizons, the level of independence of training and test data for this particular ensemble model is still just a little under 98%.

Thus, it is reasonable to suggest that data that is used in both training and test datasets may still be considered to be independent, and data may be conserved by recycling it.

The difference between the durability and overall results is composed of the results of purely seen and purely unseen tests. For a test horizon to be purely seen, it has to have been included in the training set of *each* member ANN of the ensemble. Clearly, this is almost never going to be the case, and so the influence of purely seen data is virtually zero, and can be disregarded. Thus, the overall results are composed of the results of tests with green and unseen data. Since the durability and overall results are virtually the same, then the reliability results must also be the same, and hence, results of testing may be described in terms of reliability, durability or overall, with no loss of precision.

### 6.2.2 ANN Ensemble Soil Water Retention Continuous PTF Results

So far, the tests performed have had the aim of optimising certain parameters in isolation. It now remains to pool together all of the knowledge gained and build a model using all of the optimised parameters together, to determine if results may be improved overall. However, so far it has been assumed that all parameters are independent from one another, and that they may all be assessed in isolation. This may not be the case with respect to the number of soil horizons to use per ANN and the number of ANN members to use in the ensemble. Although it may be difficult to determine the degree of interdependence between these two parameters, it is possible to measure them together empirically, and compare the results.

To achieve this, a program was constructed containing the following steps:

1. Data is extracted randomly from the HYPRES dataset by the Bootstrap method, with the same number of soil horizons from each texture class. This process is repeated 10 times, giving 10 sets of training data.
2. Each of the datasets is pre-processed by Equation 6.1b.

3. The 10 training sets are presented to 10 ANNs, which are trained to optimisation with the BR algorithm. To increase the variance of the individual ANNs (and therefore decrease the statistical bias), the initial random weights and biases from which each ANN is trained is varied from ANN to ANN.
4. The 10 optimised ANNs are tested with all the available data, determining the results (RMSE and ME, see Section 2.5) of the individual ANNs for tests with both seen and unseen data, and overall.
5. The predictions of the individual ANNs are combined to produce results (RMSE and ME) for ensembles of 1 to 10 ANN members.
6. Steps 1 to 5 are repeated 20 times.
7. The best 10 models are averaged to provide smoothing.
8. Steps 1 to 7 are repeated for 5, 10, 15, 25, 35 and 45 soil horizons from each texture class.
9. The results are plotted as (RMSE or ME) v (number of ensemble members or number of soil horizons per ANN member), each of which is detailed by texture class, and overall. This gives four sets of six graphs, Figures 6.2b-e.

Figure 6.2b is a plot of RMSE *versus* number of ANN ensemble members (from 1 to 10, in integer increments). Each graph has six plots, corresponding to 5, 10, 15, 25, 35 and 45 soil horizons per texture class. Figure 6.2c is a plot of RMSE *versus* number of soil horizons per texture class. Each graph has ten plots, corresponding to ensembles that have from 1 to 10 ANN members. Figures 6.2d and 6.2e are the results of testing with respect to ME, rather than RMSE.

Figure 6.2b shows that the ensemble method provides a more robust method of modelling soil water retention than single ANN methods, with the RMSE decreasing when the number of ANN members is increased from one to two, in almost every case. It can also be seen that, as the number of ANN members is increased, there is a decrease in the RMSE, reaching steady-state when the number of members is ten, regardless of how many soil horizons per texture class are used for the model. It is also interesting to see the results for just five soil horizons per texture class as the number of ANN ensemble members is increased. For a single ANN, the results are very poor, as expected, since the ANN has been presented with insufficient data to express the complexity of $\theta_{FC}$. As the number of ensemble members is increased, the

RMSE drops significantly and, when the number of members is increased to ten, the overall RMSE falls to a rather respectable 0.077.

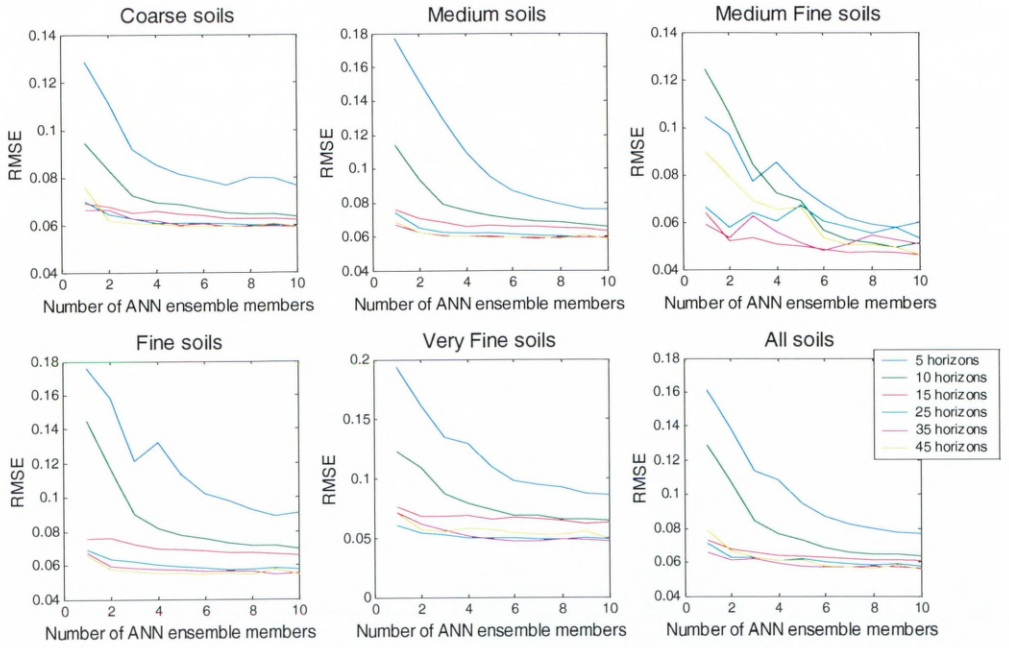



*Figure 6.2b: RMSE* versus *number of ANN ensemble members, for ANNs having between 5 and 45 soil horizons per texture class*

Figure 6.2c shows that increasing the number of soil horizons per texture class results in a decrease in the RMSE, reaching steady-state when the number of soil horizons per texture class is 25. For models having more than this amount of data, the improvements to the RMSE are minimal. In addition, note the graph for MF soils. When the number of ensemble members is low, the RMSE results show considerable variability. However, as the number of members is increased, the variability is significantly decreased, leaving a (relatively) smooth transition from few to many soil horizons per texture class. Thus, improvement of the overall RMSE is not the only advantage to be found when using ensembles.

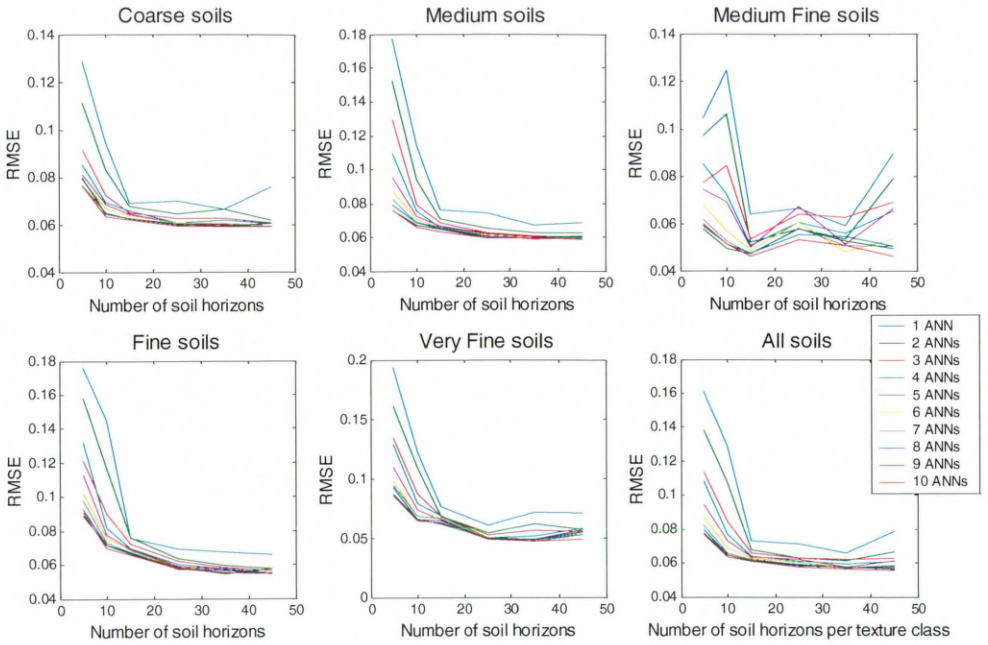



*Figure 6.2c: RMSE* versus *number of soil horizons per texture class, for ensembles having 1 to 10 ANN members*

Figure 6.2d shows that an increase in the number of ensemble members produces little or no improvement in the ME. Recall that the ME is a measure of the statistical bias of the model, and that the effect of combining estimators is to effect a reduction in the variance, but have no effect on the bias (see the bias/variance trade-off, Section 4.2.1). This result, then, is to be expected. However, note that as the number of soil horizons per texture class is increased, there is a general improvement in the bias. Note also the scale of the plots – the MEs are an order of magnitude smaller than the RMSEs, indicating that the bias in the model is small. When ANNs have greater than about 15 soil horizons per texture class, the ME is very small and very stable.

Figure 6.2e shows that as the number of soil horizons per texture class is increased, the ME tends to steady-state at ME $\approx$ 0.

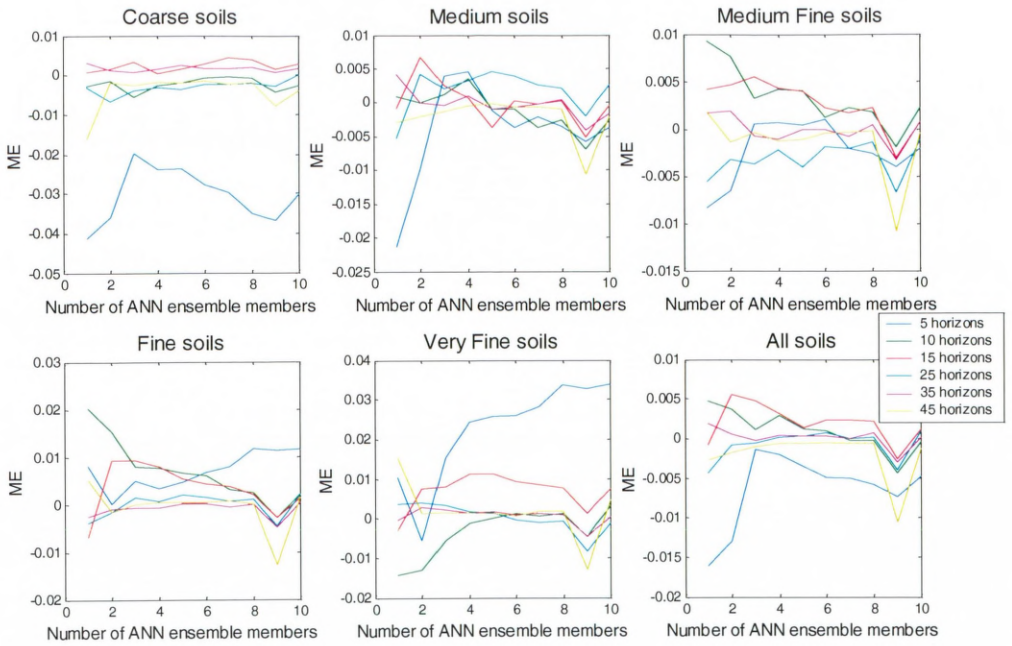



*Figure 6.2d: ME* versus *number of ANN ensemble members, for ANNs having between 5 and 45 soil horizons per texture class*

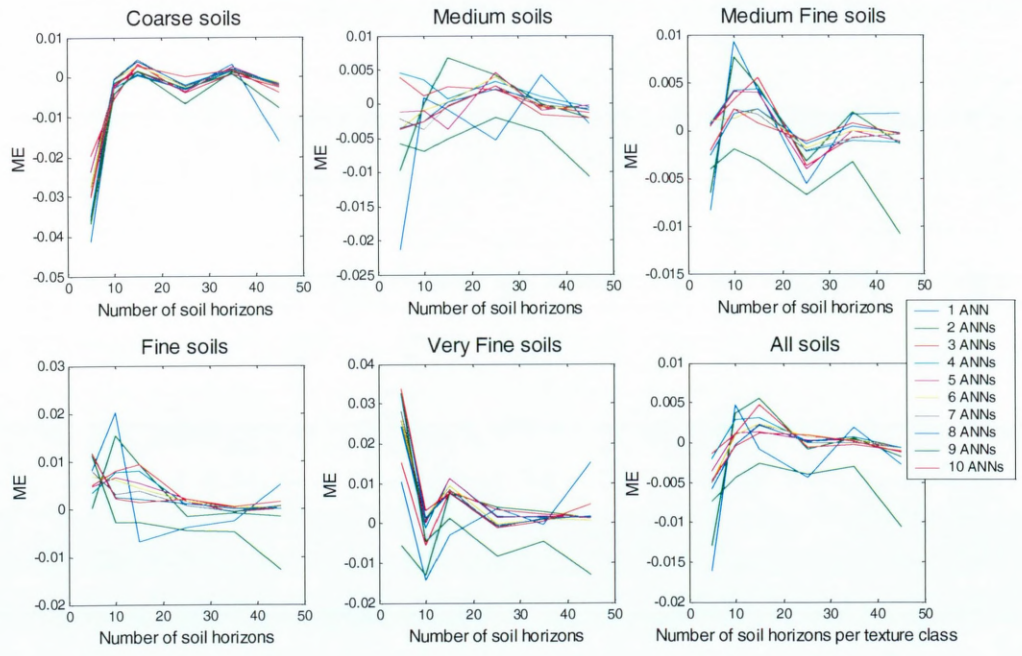



*Figure 6.2e: ME* versus *number of soil horizons per texture class, for ensembles having 1 to 10 ANN members*

The overall conclusion that may be established from Figures 6.2a-d is that modelling $\theta_{FC}$ with an ensemble of 10 members, each training 25 soil horizons per texture class will produce a stable, reliable and robust PTF model. It should be noted that, as this is what was established from Figures 6.1e and 6.1f, it may be reasonable to assume that the number of soil horizons per texture class and number of ensemble members to employ may be determined independently from each other.

Now that all parameters (ANN architecture, number of soil horizons per ANN, number of ensemble members, re-scaling algorithm and training algorithm) have been determined, it is time to construct what may be termed the 'final' optimised ANN ensemble. The procedural basis follows the following steps:

1. Data is extracted randomly from the HYPRES dataset by the Bootstrap method, with 25 soil horizons from each texture class. This process is repeated 20 times to give 20 sets of training data.
2. Each of the datasets is pre-processed by Equation 6.1b.
3. The 20 training sets are presented to 20 ANNs, which are trained to optimisation with the BR algorithm. To increase the variance of the individual ANNs (and therefore decrease the statistical bias), the initial random weights and biases from which each ANN is trained is varied from ANN to ANN.
4. The 20 optimised ANNs are tested with all the available data, determining the results (RMSE and ME) of the individual ANNs for tests with both seen and unseen data, and overall.
5. The best 10 ANNs are chosen to build the ensemble, based on the RMSE of tests of the unseen data for each of the individual ANNs.
6. The predictions are combined to produce ensemble predictions, and the RMSE and ME of the ensemble calculated.
7. Steps 1 to 7 are repeated 10 times to produce 10 ensemble models of $\theta_{FC}$.
8. The RMSE and ME results of the 10 ensemble models are averaged to effect smoothing.

At step 5, based on a given criterion, ten ANNs are chosen with which to build the ensemble, and the remaining ten are discarded. Early tests showed that approximately one quarter to one third of candidate ANNs 'fail', and thus, training 20

and discarding 10 ensure that the remaining 10 are viable ANNs that have reached optimisation and will generalise well.

At step 6, the individual ANN predictions are combined to produce ensemble predictions. In Section 4.2.3, some of the more popular methods of combining estimators were discussed, and in this research, I employed the method of weighted averaging. The expression for the weighted average $\bar{X}$ is given by:

$$\bar{X} = \frac{\sum_{i=1}^{N} w_i . p_i}{\sum_{i=1}^{N} w_i} \tag{6.2a}$$

where $p_1, p_2, \ldots, p_i, \ldots, p_N$, ($N$ is the number of ensemble members) are several independent, unbiased estimates of (in this case) $\theta_{FC}$, and $w_i$ is the frequency (or weight) of element $p_i$. However, this method of weighted averaging give a greater importance to elements having a larger weight.

In the ensemble method, the ANN that has trained to the *smallest* error goal (see Section 3.5.2) is likely to generalise better, and thus, ANNs should be weighted by error goal at optimisation, weighted so that the smaller error goal is given more weight. This is achieved by:

$$\bar{X} = \frac{\sum_{i=1}^{N} \left( \left( \left( \sum_{i=1}^{N} SSE_i \right) - SSE_i \right) . p_i \right)}{(N-1) . \sum_{i=1}^{N} SSE_i} \tag{6.2b}$$

where $p_1, p_2, \ldots, p_i, \ldots, p_N$ are several independent, unbiased estimates of $\theta_{FC}$, $SSE_i$ is the sum squared error (the error goal or weight) of the $i$th ANN, and $N$ is the number of ANNs that make up the ensemble, in this case, ten.

Table 6.2a shows the RMSE results (all tests) of each of the ten ensembles, denoted HCO_Ens_1, HCO_Ens_2, ..., HCO_Ens_10, shown to four decimal places, and detailed by texture class and overall, where HCO denotes that the model is a HYPRES Continuous PTF.

| | RMSE | | | | | |
|---|---|---|---|---|---|---|
| | C | M | MF | F | VF | ALL |
| HCO_Ens_1 | 0.0583 | 0.0591 | 0.0425 | 0.0587 | 0.0504 | 0.0555 |
| HCO_Ens_2 | 0.0592 | 0.0595 | 0.0411 | 0.0544 | 0.0427 | 0.0546 |
| HCO_Ens_3 | 0.0589 | 0.0587 | 0.0420 | 0.0563 | 0.0444 | 0.0547 |
| HCO_Ens_4 | 0.0586 | 0.0590 | 0.0440 | 0.0550 | 0.0444 | 0.0548 |
| HCO_Ens_5 | 0.0596 | 0.0583 | 0.0422 | 0.0599 | 0.0502 | 0.0557 |
| HCO_Ens_6 | 0.0582 | 0.0582 | 0.0420 | 0.0557 | 0.0458 | 0.0543 |
| HCO_Ens_7 | 0.0586 | 0.0587 | 0.0430 | 0.0562 | 0.0504 | 0.0550 |
| HCO_Ens_8 | 0.0593 | 0.0597 | 0.0435 | 0.0596 | 0.0478 | 0.0561 |
| HCO_Ens_9 | 0.0592 | 0.0594 | 0.0434 | 0.0592 | 0.0507 | 0.0561 |
| HCO_Ens_10 | 0.0593 | 0.0579 | 0.0407 | 0.0553 | 0.0421 | 0.0541 |

*Table 6.2a: RMSEs of 10 ensembles, detailed by texture class and overall*

In Section 6.1.3 it was established that, for the amount of data that is available for these investigations, the results will probably only be reliable to two decimal places. By quoting to four decimal places, I am not claiming that these results are correct to four decimal places, but rather I wish to draw attention to the remarkable stability of the ensembles, each of them varying little in terms of their predictive abilities, overall and for each texture class.

Table 6.2b shows the ME results (all tests) of each of the ten ensembles, shown to four decimal places, and detailed by texture class and overall.

| | ME | | | | | |
|---|---|---|---|---|---|---|
| | C | M | MF | F | VF | ALL |
| HCO_Ens_1 | -0.0012 | 0.0075 | 0.0032 | 0.0021 | -0.0074 | 0.0030 |
| HCO_Ens_2 | 0.0061 | -0.0065 | -0.0009 | -0.0023 | -0.0002 | -0.0015 |
| HCO_Ens_3 | 0.0062 | 0.0037 | 0.0012 | 0.0011 | -0.0020 | 0.0031 |
| HCO_Ens_4 | 0.0032 | -0.0054 | 0.0012 | -0.0039 | -0.0020 | -0.0017 |
| HCO_Ens_5 | -0.0004 | 0.0015 | 0.0001 | -0.0029 | -0.0003 | -0.0001 |
| HCO_Ens_6 | 0.0068 | -0.0064 | 0.0007 | -0.0002 | -0.0014 | -0.0006 |
| HCO_Ens_7 | -0.0023 | 0.0016 | 0.0009 | -0.0057 | 0.0015 | -0.0008 |
| HCO_Ens_8 | 0.0021 | -0.0007 | 0.0019 | 0.0051 | 0.0035 | 0.0017 |
| HCO_Ens_9 | 0.0072 | 0.0056 | 0.0047 | -0.0025 | -0.0044 | 0.0039 |
| HCO_Ens_10 | -0.0033 | $-5.18\times10^{-5}$ | 0.0008 | -0.0076 | -0.0008 | -0.0020 |

*Table 6.2b: MEs of 10 ensembles, detailed by texture class and overall*

Here, all MEs are at least one order of magnitude smaller than the corresponding RMSE results, indicating that the bias in each of the models is very small. It is also encouraging to note that the sign of the ME is not consistent, in other words, $\theta_{FC}$ is neither consistently under- nor over-estimated. This is in contrast to results published by Nemes *et al.* (2003) who found that water retention was always under-estimated in their models, indicating a significant bias towards datasets having small %OM. Although these researchers used the ensemble method as outlined by Schaap *et al.* (2001), they did not take care to reduce bias in the dataset before training (bias/variance trade-off, see Section 4.2.1), and this may account for their findings.

Table 6.2c shows the average results, both RMSE and ME, of the ten ensembles, detailed by texture class and overall.

| | C | M | MF | F | VF | ALL |
|---|---|---|---|---|---|---|
| **Number of tests** | 630 | 952 | 574 | 516 | 107 | 2779 |
| **RMSE** | 0.059 | 0.059 | 0.042 | 0.057 | 0.047 | 0.055 |
| **ME** | 0.0024 | $7x10^{-5}$ | 0.0014 | -0.0017 | -0.0014 | 0.0005 |

*Table 6.2c: RMSE and ME results, averaged over 10 ensemble models*

The RMSEs for C, M and F texture classes are very similar, whilst the RMSEs of MF and VF are significantly better. Recall, however, that for testing to be valid at the two decimal point level, there should be ~150-200 tests or more. There are only 107 VF soil horizons available for testing, thus, the RMSE for predictions of VF soils has not necessarily reached steady-state. Nevertheless, if more VF data were available, it is not expected that further testing would have a significant effect on the overall result.

Note again that all MEs are at least one order of magnitude smaller than the corresponding RMSE results.

Since good predictions of soil hydraulic properties rather than direct measurement may be adequate for many applications, a definition of the adjective 'good' remains to be defined. As a starting point, the reader is directed to Table 5.1a, which details a selection of water retention PTFs from the literature. In these examples, the RMSE of volumetric water contents ranges from 0.02 to 0.11, and thus, for any PTF to be described as 'good', it must have comparable RMSEs (or better) to these. The ensemble method presented here, utilising selected HYPRES data, yields an overall

RMSE of 0.055 (Table 6.2c). This value lies centrally within the range of 'good' PTFs, and it may therefore be concluded that the ensemble method produces PTFs of adequate quality. It should be noted, however, that the aim of this research was not to produce a PTF with the lowest possible RMSE, but rather to effect a trade-off between the *amount* of data and the *value* of that data in order to produce a PTF of acceptable quality. In Section 4.1, it was ascertained that researchers typically take ~2000 soil horizons and use approximately 2/3 and 1/3 of these for training and testing, respectively. The research detailed here shows that, using the ensemble method, significantly less data than this is required in order to construct a PTF of acceptable quality. The question remaining here, is, when compared to competing methods, how much less data has been used? This question is the subject of Sections 6.2.3 and 6.2.4.

### 6.2.3 Single ANN Soil Water Retention Continuous PTF Results

To date there has not been an ANN constructed using the HYPRES dataset, so there are no current benchmarks against which the results may be compared.

Nemes *et al.* (2003) published ANN results from a dataset termed 'HUNSODA' (Hungarian UNSODA). They collected all available Hungarian soil horizons that had appropriate input and output parameters (SSC, BD and %OM, and $\theta$) to form a single dataset, which was then split into a dataset for training (HUN, 235 soil horizons) and a dataset for testing (HUNTEST, 236 soil horizons). They also extracted data from HYPRES (2464 soil horizons) and constructed a dataset of international soil data (ICO, 1347 soil horizons). From these they created a further two datasets, HYPRES + HUN (2699 soil horizons) and ICO + HUN (1582 soil horizons), giving five datasets in all for training and one dataset for testing.

Each of the training datasets was Bootstrapped in the ratio 2:1 for training and validation data, and the ANNs were trained. This procedure was repeated 50 times and predictions were averaged (i.e. an ensemble method using an averaging technique to combine), and RMSEs and MEs calculated.

It should be noted here that all the ANNs were tested with Hungarian soil horizons only – although a HYPRES ANN was constructed, it has not been tested with HYPRES data, only with Hungarian data. Since my methods and test data are different to those of Nemes *et al.* (2003), the results may not be directly compared.

Although these researchers used an ensemble whose results may not be compared with those of my ensemble, they have recorded the only published results of an ANN-PTF constructed from HYPRES data.

However, constructing a single ANN using similar techniques (in particular, using the same amount and distribution of data) to those of Nemes *et al.* (2003), a comparison of these results with those from my ensemble method may be performed. To achieve this, the same number of soil horizons were extracted from HYPRES as Nemes *et al.* (2003) used for HUN. However, although in this thesis I have shown that it is advantageous to model with equal numbers of soil horizons from each texture class to reduce bias (see Section 4.2.1), Nemes *et al.* (2003) did not do this, thus HYPRES training data has been extracted proportionally to the amount of data available. This resulted in a dataset that contains $\frac{630}{2779}\times 235$ C soil horizons, $\frac{952}{2779}\times 235$ M soil horizons, etc..

This data was trained to optimisation in a single ANN, utilising all of the lessons learnt from past experience (see Sections 6.1.1 – 6.1.7), and tested with all the available data. The process was repeated 20 times to facilitate smoothing of the RMSE and ME results. In addition, single ANNs with two, three, four and five times this amount of data were trained (termed the x1, x2, x3, x4 and x5 datasets), and tested with all available data. The x1 and the x5 RMSE and ME results are detailed in Table 6.2d, and the RI of the x5 over the x1 RMSE results are shown in Table 6.2e.

| | | C | M | MF | F | VF | ALL |
|---|---|---|---|---|---|---|---|
| x1 | **N° of soil horizons (training)** | 53 | 81 | 49 | 44 | 9 | 236 |
| | **RMSE** | 0.063 | 0.062 | 0.044 | 0.057 | 0.060 | 0.058 |
| | **ME** | -0.0061 | -0.0033 | 0.0006 | -0.0034 | -0.0104 | -0.0034 |
| x5 | **N° of soil horizons (training)** | 265 | 405 | 245 | 220 | 45 | 1180 |
| | **RMSE** | 0.059 | 0.058 | 0.040 | 0.055 | 0.048 | 0.054 |
| | **ME** | 0.0021 | -0.0020 | -0.0011 | -0.0014 | -0.0033 | -0.0008 |

*Table 6.2d: RMSE and ME results, averaged over 20 single ANN models*

| | C | M | MF | F | VF | ALL |
|---|---|---|---|---|---|---|
| **RI (%, x5 over x1)** | 6 | 7 | 9 | 3.5 | 20 | 7 |

*Table 6.2e: RI of the RMSE results of dataset x5 over x1*

Increasing the number of soil horizons for training did not make significant improvements to the RMSEs of any texture class, with the exception of VF soils. It would be expected that the RMSE of this particular texture class would be improved on increasing the amount of training data, since there were only nine soil horizons in the x1 training set, perhaps an amount that is inadequate to capture the complexities of $\theta_{FC}$. Clearly, increasing the amount of training data five-fold yields only minor RMSE improvements to all except the VF texture class, less than 9% RI in all cases, and these small improvements are not worth all the extra work and computing time.

What is required then, is to increase the quality in the modelling procedures and the data, not the quantity.

### 6.2.4 Continuous PTF Results Comparison and Discussion

Pooling all the results together, we may compare the ensemble and single ANN results, Table 6.2f. For this, I chose HCO_Ens_10, since this had the smallest RMSE(ALL) compared will all of the other ensembles.

| | | C | M | MF | F | VF | ALL |
|---|---|---|---|---|---|---|---|
| RMSE | HCO_Ens_10 | 0.059 | 0.058 | 0.041 | 0.055 | 0.042 | 0.054 |
| | x1 | 0.063 | 0.062 | 0.044 | 0.057 | 0.060 | 0.058 |
| | x5 | 0.059 | 0.058 | 0.040 | 0.055 | 0.048 | 0.054 |
| ME | HCO_Ens_10 | -0.003 | $-5x10^{-5}$ | 0.001 | -0.008 | -0.001 | -0.002 |
| | x1 | -0.006 | -0.003 | 0.001 | -0.003 | -0.010 | -0.003 |
| | x5 | 0.002 | -0.002 | -0.001 | -0.001 | -0.003 | -0.001 |
| N° of tests (all) | | 630 | 952 | 574 | 516 | 107 | 2779 |
| N° of training horizons | HCO_Ens_10, per ANN | 25 | 25 | 25 | 25 | 25 | 125 |
| | x1 | 53 | 81 | 49 | 44 | 9 | 236 |
| | x5 | 265 | 405 | 245 | 220 | 45 | 1180 |

*Table 6.2f: Comparison of results of ensemble model HCO_Ens_10 and single ANN models x1 and x5*

Overall results (RMSE) from HCO_Ens_10 and the single ANN approach are equivalent when the single ANN requires ~1000 soil horizons (model x5). For texture classes C, M, MF and F, HCO_Ens_10 and x5 produce the same results, however, VF testing for model x5 produces results that are significantly worse that those for

HCO_Ens_10. Thus, it appears that more data than is available to train x5 is required in order to generate results that are at least equivalent to results of HCO_Ens_10 for all texture classes.

Figures 6.2f and 6.2g show how the ensemble-predicted and single ANN predicted $\theta_{FC}$ values correlate to the expected values, for HCO_Ens_10 and the best single ANN model (determined by overall RMSE) of dataset x5, respectively.

There is quite good agreement between expected and predicted values of $\theta_{FC}$ for all texture classes and overall, for both the ensemble and single ANN methods, with relatively tight bunching about the expected 1:1 correspondence lines. There does not appear to be consistent under or overestimations for any of the texture classes or overall.

Comparing Figures 6.2f and 6.2g, the correspondence between expected and predicted values of $\theta_{FC}$ for the single ANN is bunched a little tighter than its ensemble counterpart for C soils, whilst the correspondence for the ensemble model of texture classes M, MF and VF are bunched a little tighter than in the single ANN model.

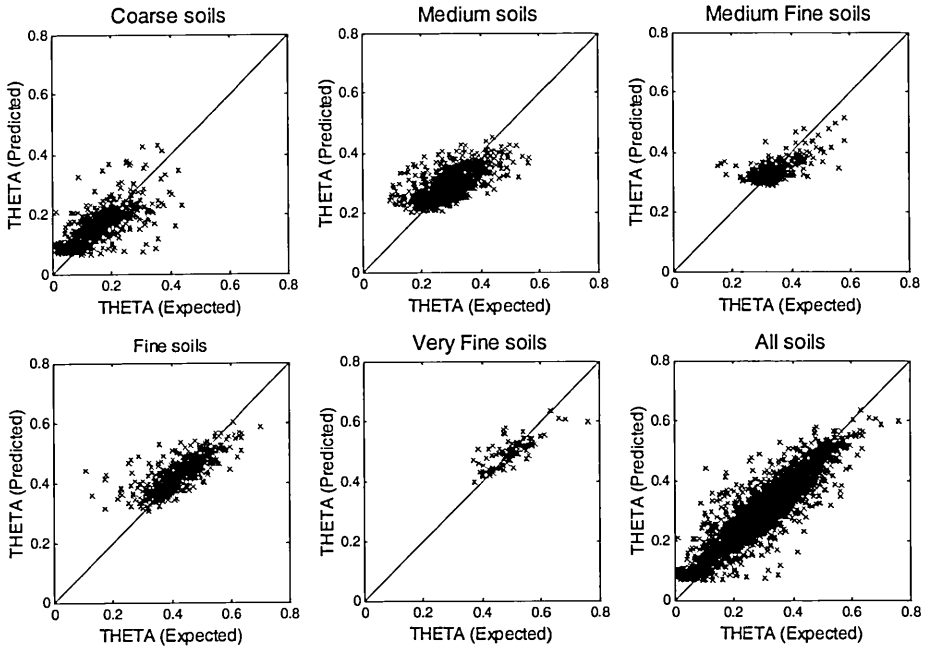



*Figure 6.2f: Comparison of expected and ensemble-predicted $\theta_{FC}$ for ensemble model HCO_Ens_10, detailed by texture class and overall*

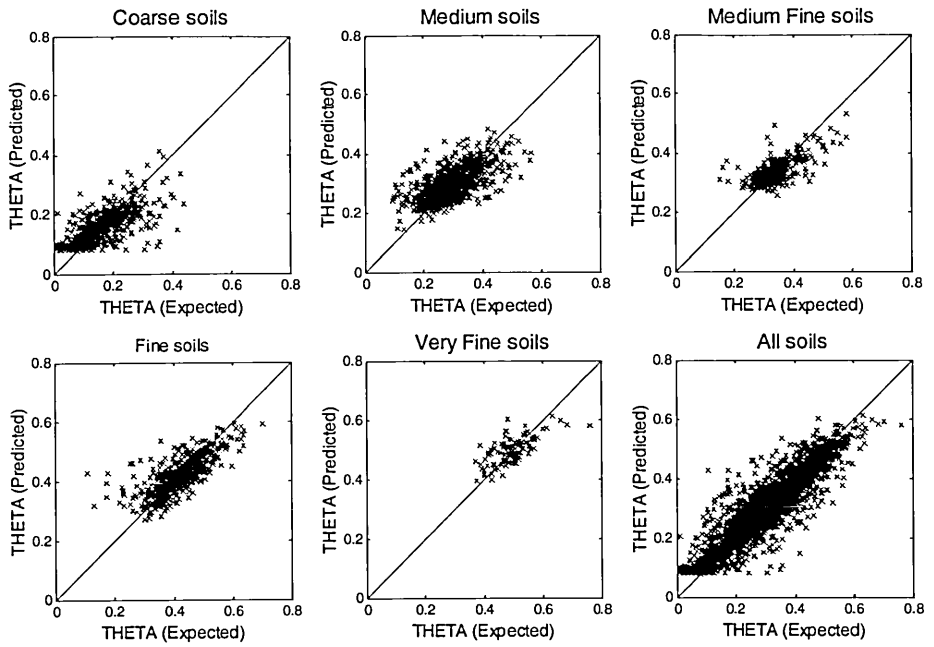



*Figure 6.2g: Comparison of expected and single ANN predicted $\theta_{FC}$ for single ANN model x5, detailed by texture class and overall*

Consider that ~1000 soil horizons were used for the training regime of x5 (excluding a validation set), and that ~150-200 soil horizons are required from each texture class for testing. This means that a single ANN method, such as that outlined here, would require more than 2000 soil horizons in order for both training and testing to be stringent, and the RMSE results to be the same as HCO_Ens_10. In my model of the single ANN method, the BR algorithm was utilised for training, whilst Nemes *et al.* (2003) used the LM algorithm. Recall that one important difference is that the LM algorithm requires a validation set, whilst the BR algorithm does not. Thus, a single ANN method employing the LM algorithm would require that a further 500-1000 soil samples be used for validation. Overall, a single ANN using the LM algorithm would require ~3000 soil horizons or more to ensure that both training and testing were conducted stringently.

Conversely, the ensemble method requires ~200 soil samples from each texture class for testing, ~1000 overall. Since this method samples with replacement from the dataset for training, as long as there is sufficient data for testing, no additional data

need be set aside for training. Neither is there a need for a validation set when the BR algorithm is used.

Thus, an ensemble requires approximately one-third of the amount of soil horizons to provide equivalent results to the single ANN method described here. In a situation where data is in short supply, this is a crucial result.

Nemes *et al.* (2003) demonstrated that a PTF constructed from local data performs better than one built from data spread from a wider area (which is why they did not test the HYPRES ANN-PTF with a HYPRES test set). In the light of this result, since local data is less bountiful, data conservation becomes a critical factor in ANN-PTF construction, and ensembles appear to be the answer to the problem since they allow data to be recycled.

### 6.2.5 Sensitivity Analyses of the Ensemble

So far, all the research detailed herein has been focussed towards extracting the maximum information from the available data. However, is there anything that can be done to improve the quality of the data? The way to achieve this is to investigate the source of errors in individual data samples and find some way of correcting those errors. HYPRES data has been sampled, measured and recorded using varying protocols that may have led to a range in the quality of data samples. Clearly, finding the source of all the errors is not practicable. However, is it possible to determine where the major errors typically lie in order to minimise them when new field measurements are taken? Sensitivity analyses of the ensemble model can help inform field scientists which particular parameters are of most importance and should be measured with most care and diligence.

Each of the five input parameters has been omitted in turn from the ensemble model, also SSC was omitted, and BD & %OM, and then each of these re-trained and re-tested in precisely the same way as in the ensemble method. This procedure is repeated 10 times, and the results averaged. The RMSE results of testing are shown in Table 6.2g, and detailed by texture class. The results are also shown graphically in Figure 6.2h. Additionally, the RMSE results of the ensemble with no parameter omissions are repeated for comparative purposes (results averaged over 10 procedural repeats). The ME results varied very little from model to model and are not shown here.

| Input parameters used | | | | | RMSE | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| % Cl | % Si | % Sa | BD | %OM | C | M | MF | F | VF | ALL |
| ✓ | ✓ | ✓ | ✓ | ✓ | 0.059 | 0.059 | 0.042 | 0.057 | 0.047 | 0.055 |
| | ✓ | ✓ | ✓ | ✓ | 0.059 | 0.059 | 0.042 | 0.057 | 0.043 | 0.055 |
| ✓ | | ✓ | ✓ | ✓ | 0.059 | 0.059 | 0.044 | 0.055 | 0.043 | 0.055 |
| ✓ | ✓ | | ✓ | ✓ | 0.059 | 0.059 | 0.043 | 0.055 | 0.043 | 0.055 |
| ✓ | ✓ | ✓ | | ✓ | 0.061 | 0.063 | 0.045 | 0.071 | 0.061 | 0.061 |
| ✓ | ✓ | ✓ | ✓ | | 0.060 | 0.060 | 0.044 | 0.054 | 0.043 | 0.056 |
| ✓ | ✓ | ✓ | | | 0.067 | 0.070 | 0.049 | 0.076 | 0.066 | 0.066 |
| | | | ✓ | ✓ | 0.149 | 0.080 | 0.066 | 0.077 | 0.079 | 0.097 |

*Table 6.2g: RMSEs of ensemble sensitivity analyses, averaged over 10 models*

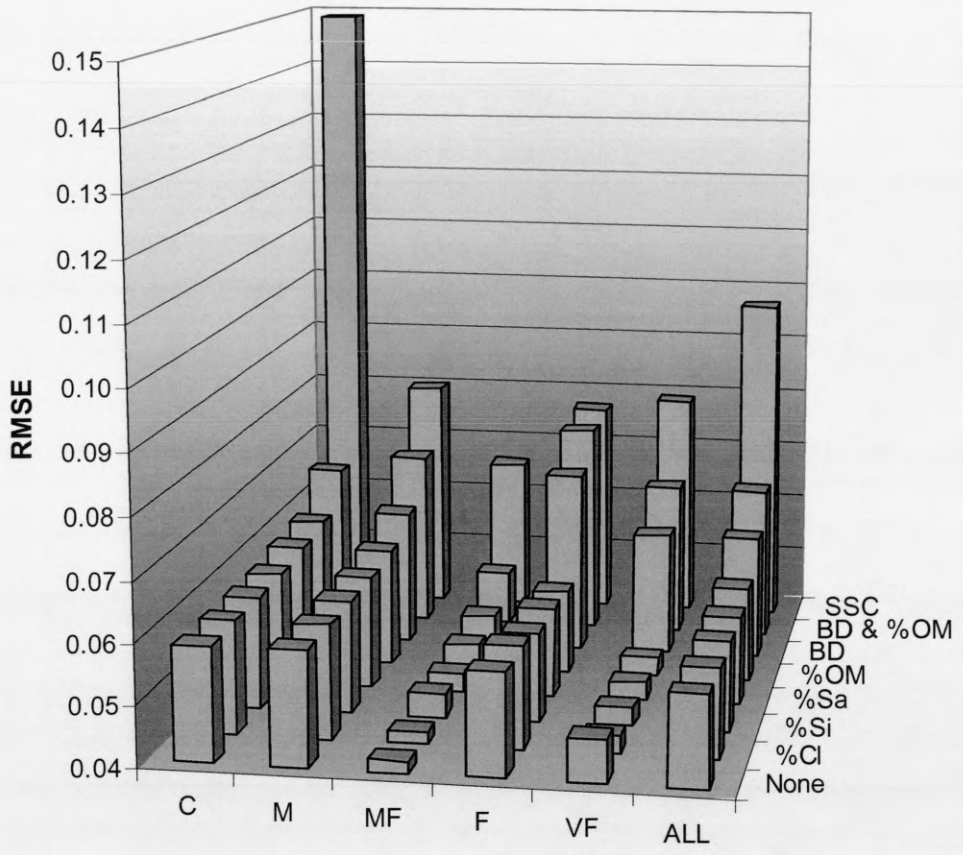



*Figure 6.2h: RMSEs of ensemble sensitivity analyses, averaged over 10 models, detailed by texture class and overall. Omitted parameter(s) are shown on the z-axis*

**Comparing ensemble with ensemble minus %Cl:**

Removing clay from the model improves the ability of the ensemble model to predict VF soils. Since clay is the major component of the VF soil texture class (see the soil textural triangle, Figure 2.1b), and adsorption of water is the mechanism causing the strong retention of water by clay soils, removing it from the model should decrease its predictive abilities. Recall that ~150-200 soil horizons are required for stringent testing at the two decimal point level (see Section 6.1.3), and also that there are only 107 soil horizons in the VF texture class. Perhaps these results have not yet reached steady-state.

**Comparing ensemble with ensemble minus %Si:**

On comparing the five-parameter ensemble with the ensemble minus %Si, the predictive ability for the C and M texture classes is unaffected by the removal of %Si from the model. Results in the MF category worsened a little with the removal of %Si, and this is also expected, since silt is the major component of MF (Figure 2.1b). However, F and VF soils were predicted with slightly greater durability by the four-input ensemble, and this was an unexpected result, since %Si is a major component of these texture classes (Figure 2.1b).

**Comparing ensemble with ensemble minus %Sa:**

Omission of %Sa from the model does not affect the predictive ability of the ensemble with respect to C and M soils. At first glance, this seems counter-intuitive, since sand content is a major component of these texture classes. However, since $\theta_{FC}$ is more influenced by the clay particles that lodge in the pore spaces than by the sand particles that define the sizes and distribution of the pore spaces, on balance it is perhaps not surprising that sandy soils are not affected by the omission of %Sa from the model. MF results worsened slightly, as expected, since sand content is a minor component of MF soils, and the RMSEs of F and VF soils improved. This is also expected, since sand is a minor component of these texture classes, and therefore any sand content may only act as noise in the model. Additionally, the variability of sand in the F and VF texture classes is very high, and removing this parameter may act to remove uncertainty from the model for these texture classes.

**Comparing ensemble with ensemble minus BD:**

Omitting BD from the model results in a decrease in predictive ability for all texture classes. Since many previous publications have indicated that BD is an important parameter in PTFs (Rawls *et al.* 1982, Aina and Periaswamy 1985, Rajkai and Várallyay 1992, Bruand *et al.* 1996, Wösten *et al.* 1999), this is a wholly expected result. Moreover, BD appears to be increasingly important for soils having increasing clay content.

**Comparing ensemble with ensemble minus %OM:**

Predictions for C, M and MF soils worsened slightly, whilst predictions of F and VF soils improved a little. Bloemen (1980) found a correlation between BD and %OM, and indicated that BD effectively substituted %OM. The results here do not concur with those findings. In fact, the results appear to indicate that %OM becomes less important with increasing clay content. It would be expected that omission of %OM from the model would result in a significant decrease in the predictive abilities of the model, since the degree of adsorption between organic matter and water is very high. It is, therefore, a little surprising to find that the omission of %OM has only a small effect on the predictive abilities of the ensemble model.

Overall, considering the effects of omitting a single parameter, BD appears to be very important for the predictive ability of the ensemble for all texture classes, while no other single parameter had a significant effect on any texture class.

**Comparing ensemble with ensemble minus BD & %OM:**

Essentially, this ensemble model utilises only particle sizes. Not surprisingly, the predictive abilities of the model are reduced significantly for all texture classes, but predictions remain reasonably respectable. Clearly, the leading factor is the omission of the BD parameter.

**Comparing ensemble with ensemble minus SSC:**

This is a model for which only BD and %OM are used. The RMSEs here are very much worse for all texture classes. The overall conclusion made by previous investigators (see Section 4.1) was that when the number of inputs to an ANN is

greater than three, ANNs usually perform better than regression techniques. This suggests that ANNs with just two inputs have decreased predictive abilities, compared to those having more inputs than two, and this is exactly what is seen here. However, the results are still relatively respectable. Omission of all textural data deprives the model of information regarding the size and distribution of the components that constitute the structure of the soil and the extent of the pore spaces. On omitting SSC it would be expected that the model would completely fail to capture the necessary complexity of $\theta_{FC}$, but this is clearly not the case.

**Comparing results by texture class:**

RMSE results indicate that MF soils were predicted with greater accuracy than any of the other texture classes, regardless of parameter omissions. VF soils were also predicted more accurately than the other texture classes, regardless of parameter omissions, however, since there were only 107 soil horizons in this texture class, it is not clear whether the results have reached steady-state. C, M and F soils were predicted with similar accuracies. In terms of single parameter omissions, only BD has a significant effect, worsening the predictive ability of the ensemble. Omitting BD and %OM, the predictions deteriorate, and the deterioration increases with the omission of SSC. This is the case with all texture classes.

Additionally, Schaap and Bouten (1996) found that soil water retention is unrelated to particle size distribution, BD and %OM for sandy soils (mostly C, but also M). These results back up those findings, at least at $\theta_{FC}$, with the exception that BD appears to be an important parameter in ensemble models, for all texture classes.

So, overall, as a single parameter, only BD has any significant effect on the sensitivity of the ensemble's ability to predict $\theta_{FC}$, but what effect does BD have on the prediction *when misread*? By this, I mean, consider what effect the BD has on $\theta_{FC}$ when it is misread to be, say, 10% (or so) higher or lower than the actual value. To visualise this, I chose five M soil horizons at random from HYPRES and, varying the BD in increments of 0.01 in the range [0, 2], re-tested HCO_Ens_10 and plotted the results as $\theta_{FC}$ *versus* BD. The soil horizons that were chosen are shown in Table 6.2h.

| | %Cl | %Si | %Sa | BD | %OM | $\theta_{FC}$ |
|---|---|---|---|---|---|---|
| **TEST 1** | 33.6 | 49.3 | 17.1 | 1.537 | 0.94 | 0.358 |
| **TEST 2** | 9.9 | 40.6 | 49.5 | 1.812 | 0.58 | 0.210 |
| **TEST 3** | 11.4 | 40.58 | 48.02 | 1.447 | 4.08 | 0.305 |
| **TEST 4** | 25.9 | 24.81 | 49.29 | 1.408 | 2.89 | 0.360 |
| **TEST 5** | 18.4 | 13.77 | 67.83 | 1.786 | 0.19 | 0.184 |

*Table 6.2h: Five randomly selected M soil horizons*

The results of such tests are shown graphically in Figure 6.2i, additionally, the BD has been plotted against measured values of $\theta_{FC}$ for all the M soils used in the modelling process, and a linear regression line plotted.

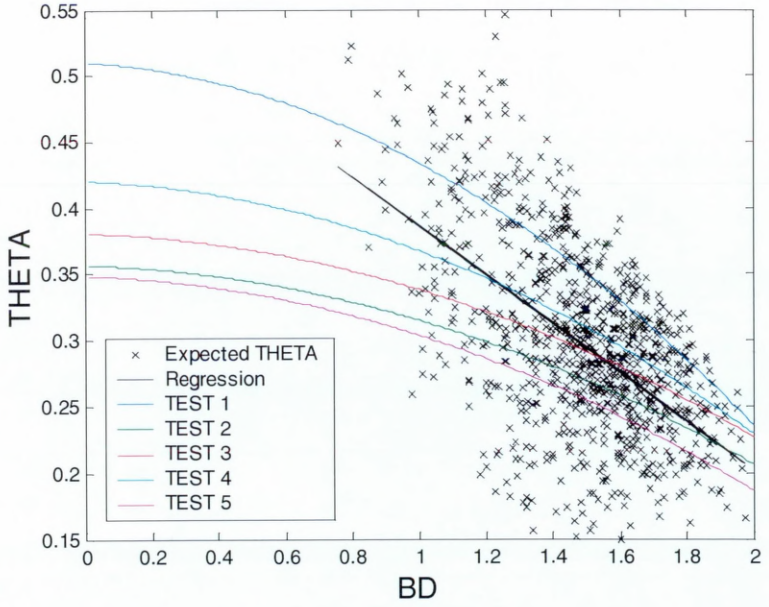



*Figure 6.2i: Effect of varying BD on five randomly chosen M soil horizons, with measured values of $\theta_{FC}$ and regression line*

It is to be expected that as the BD increases, $\theta_{FC}$ decreases, since greater values of BD indicate a greater amount of matter in the sample, and thus decreased pore space. Consequently, as the pore space is decreased, the amount of water that the soil can retain is reduced. For all M soils tested, as BD increases, $\theta_{FC}$ decreases, and the magnitude of the decrease is relatively consistent, with the exception of TEST 1, where the decrease is a little more pronounced. Additionally, the gradient of the

regression line in Figure 6.2i is not significantly different from the gradients of the five test samples shown in the range [0.8, 2].

From Tables 5.2c and 5.2e, the range of BD for M soils is [0.76, 1.97], with a mean of 1.5, and the range of $\theta_{FC}$ for M soils is [0.1, 0.57], with a mean of 0.3. The regression line on Figure 6.2i shows that one would expect that as the BD is increased $\theta_{FC}$ should decrease, and this is exactly what is seen. It should be noted that there are no soil samples that have value of BD = 0, since this would represent a soil that has zero density, i.e. a solid pore composed entirely of air, and thus, the model has no pedologic meaning for soils having BD < 0.76 (since the model would be extrapolating). Similarly, the BD of quartz is about 2.6 – 2.7, and therefore there are no soil samples having BD values close to this, since that would represent a solid block of quartz. Consequently, the model has no pedologic meaning for soils having BD > 2 (since the model would be extrapolating). As a further point of reference, water has a bulk density value of BD = 1 at standard temperature and pressure, and since most mineral soils are more dense than water, most values of BD will therefore be greater than one. Clearly, although the model may make predictions of $\theta_{FC}$ for soil horizons having values of BD < 0.76 and BD > 2, a certain amount of knowledge of soil physics is required in order to make common sense judgements and disregard these results. Models should not be asked to extrapolate beyond their range of knowledge.

This procedure has also been repeated with respect to %OM, and incremented by 0.1 in the range [0, 10]. The results of $\theta_{FC}$ *versus* %OM are shown in Figure 6.2j, additionally, the %OM has been plotted against measured values of $\theta_{FC}$ for all the M soils used in the modelling process, and a linear regression line plotted.

As %OM is increased, $\theta_{FC}$ increases for all M soils tested, although the magnitude of the effect differs from horizon to horizon. This is to be expected, since as the amount of %OM is increased, the degree of adsorption between the soil water and the organic matter is increased, thereby increasing the degree of water retention.

From Tables 5.2c and 5.2e, the range of %OM for M soils is [0, 9.4], with a mean of 1.6 (and a median of 1.2), and the range of $\theta_{FC}$ for M soils is [0.1, 0.57], with a mean of 0.3. The regression line on Figure 6.2j shows that one would expect that as

the %OM is increased $\theta_{FC}$ should increase, and this is exactly what is seen for all sample tests. It should be noted that the majority of %OM values for M soil horizons at $\theta_{FC}$ fall in the range [0, 3], and since there is much less data in the range [3, 10], it should be expected that the model produces more reliable predictions for %OM < 3. The gradient of the regression line in Figure 6.2j is not significantly different from the gradients of the five test samples shown.

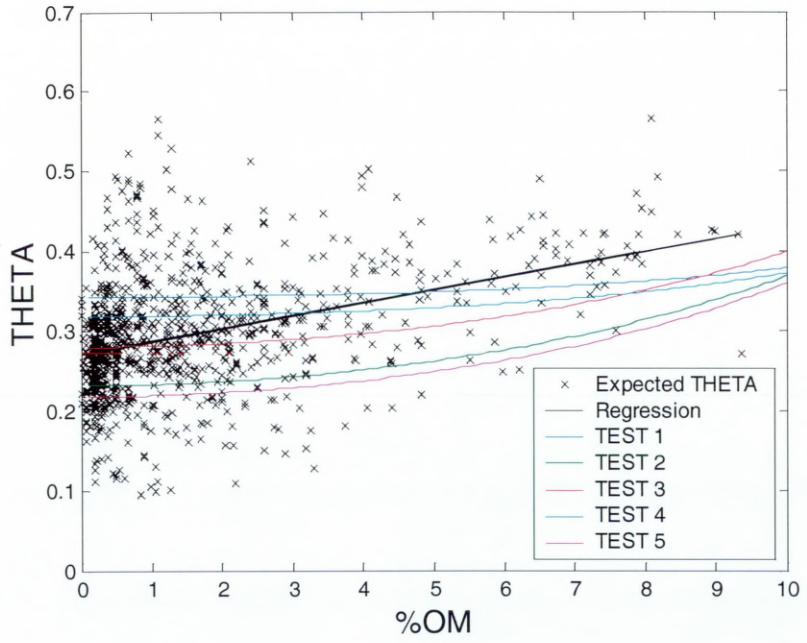



*Figure 6.2j: Effect of varying %OM on five randomly chosen M soil horizons, with measured values of $\theta_{FC}$ and regression line*

Let us suppose that TEST 5 has %OM = 0.8, and consider the effect of mistakenly entering this value into the database as, say, %OM = 8. According to HCO_Ens_10, this would result in an increase in the prediction of $\theta_{FC}$ from 0.22 to 0.30. We may also try this out on TEST 5 by supposing that BD = 0.2, and considering the effect of mistakenly entering this value into the database as, for instance, BD = 2. According to the ensemble, this would result in a decrease in the prediction of $\theta_{FC}$ from 0.35 to 0.18. These are considerable differences, and would clearly lead to significant errors in the database, and concomitant errors in models where this soil horizon is used, whether in model construction or in testing.

Thus, both BD and %OM should be evaluated with care to ensure that errors are kept to a minimum.

## 6.3 Adding Error Bars to Predictions

In (for instance) safety critical environments, such as medical or nuclear power station monitoring equipment, classification or regression performance of an ANN or ensemble is not sufficient. One of the key requirements of any statistical analysis system should be to assess its own confidence in a decision in the case of classification, and estimate probable error bars in the case of regression.

Here I present two possible ways of producing errors of individual predictions – a Bootstrapping method, and an error ensemble method. In this section, both methods will be discussed in detail, in preparation for their application to the ensemble method in the subsequent section.

### 6.3.1 Error Bars – a Bootstrapping Method

It has always been good scientific practice to take multiple readings of a given variable. The average reading will be a more accurate estimate than a single reading, since reader errors will cancel, and average to zero. This is the same principle as the ensemble method. When a single ANN is used to make predictions, confidence may not be determined for individual results. However, when a number of training sets are Bootstrapped from the available data and trained with individual ANNs, the spread of predictions (calculated as the standard deviation) will give a measure of confidence in those predictions. For this reason, this will be termed the Bootstrapping method of confidence prediction.

In terms of the ensemble method for prediction of $\theta_{FC}$, we have ten ANNs that may each make a prediction for a given soil horizon of what they believe the value of $\theta_{FC}$ is, based on the training they each received. These ten predictions will typically not be the same. Often, the expected value for $\theta_{FC}$ will lie at or near to the (weighted) average prediction, with the ten individual predictions forming a spread around the (weighted) mean. In Figure 6.3a, there are ten hypothetical individual ANN

predictions of $\theta_{FC}$ (crosses), in the range [0.26, 0.31] with a weighted mean (black triangle) of 0.285.

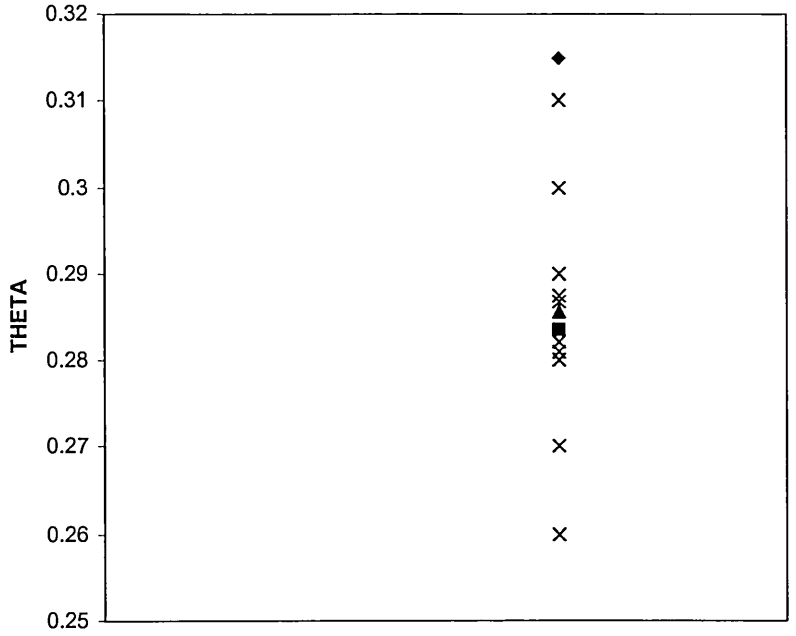



*Figure 6.3a: Ten hypothetical predicted values of $\theta_{FC}$ (crosses) with a weighted mean (black triangle), and two hypothetical measured values of $\theta_{FC}$ (black square and black diamond)*

We can see that, in this hypothetical case, the spread of individual predictions is small, indicating that the individual ANNs are in close agreement as to what the prediction should be, and therefore confidence in the ensemble prediction is high. The weighted mean of these values is 0.285, with a standard deviation of 0.014. This means that the Bootstrapped prediction of error bars for these ten predictions of $\theta_{FC}$ is 0.299 (the mean plus one standard deviation) for the upper error bar, and 0.271 (the mean minus one standard deviation) for the lower error bar. Thus, the measured value of $\theta_{FC}$ is expected to fall somewhere in the range [0.271, 0.299]. Let us consider that the black square ($\theta_{FC}$ = 0.284) is the measured value of $\theta_{FC}$. In this case, there is good agreement between the measured and ensemble-predicted values of $\theta_{FC}$, and there is high confidence in the ensemble prediction since there is a small spread of individual predictions.

The Bootstrapping method of error prediction has been used in ANN-PTF modelling by Schaap and Leij (1998a), Schaap and Leij (1998b), Schaap *et al.* (1998), Schaap *et al.* (2001) and Nemes *et al.* (2003). The latter papers illustrate improved research methods of the former (with the exception of Nemes *et al.* (2003), who used the methods detailed in Schaap *et al.* (2001)). All show that a number of individually trained ANNs may be used to make predictions of soil water retention when training data has been Bootstrapped, and that the spread of the predictions will yield a measure of confidence about them. As far as can be established, these are the only published papers that describe the ensemble method of ANN-PTF modelling. However, the crucial point of the bias/variance trade-off has not been mentioned in these papers. Soil horizons to use in the training process were Bootstrapped from the available data such that 63% of the available data was used for training, per ANN, in all cases, and 60 such ANNs were trained to predict soil water retention. Differing methods of combining the predictions have not been discussed, the authors combined by simply averaging the 60 individual predictions, and nor were any other ANN or ensemble optimisation procedures considered. In fact, the word 'ensemble' (or indeed any other term describing uses of multiple ANNs, such as 'committee', etc.) has not been mentioned in any of these papers. Although the methods used by these authors are an important step forward in the development of ANN-PTFs, I strongly suspect that they may be unaware that the methods they have used have been well described in the literature and are gaining increasing acceptance in other areas of research. The ensemble method clearly offers greater benefits to the modeller than these researchers have recognised.

### 6.3.2 Error Bars – an ANN Ensemble Method

Now assume that the black diamond in Figure 6.3a ($\theta_{FC}$ = 0.315) is the measured value of $\theta_{FC}$. The small spread of predictions in the Bootstrapping method indicates that confidence in the ensemble prediction is high. However, this is misleading, since all individual ANNs under-estimated $\theta_{FC}$. In this case, what we would hope to find is a wide spread of individual predictions to indicate a low confidence in the ensemble prediction, but the Bootstrapping method does not take account of the expected value,

only the spread of individual predictions. Thus, high confidence, in this case, is misleading.

The Bootstrapping method of confidence estimation breaks down when the member ANNs agree closely with their predictions, but consistently under- or over-estimate the expected value. In this case, the spread of individual predictions may be small, giving a high degree of confidence, and yet the ensemble prediction may be considerably different to the expected value.

Fortunately, there is another method that may be employed to estimate the spread of predictions, and involves the ensemble techniques that we have already encountered. Consider the following example of a soil horizon that is used to test the ANN ensemble HCO_Ens_10, with corresponding expected ($\theta_{FC}$) and predicted ($\hat{\theta}_{FC}$) values:

| %Cl | %Si | %Sa | BD | %OM | $\theta_{FC}$ | $\hat{\theta}_{FC}$ | $\theta_{FC}$-$\hat{\theta}_{FC}$ |
|---|---|---|---|---|---|---|---|
| 8.1 | 11.52 | 80.38 | 1.37 | 4.09 | 0.18 | 0.16 | 0.02 |

The difference between the expected and predicted values, $\theta_{FC}$-$\hat{\theta}_{FC}$, is a measure of the error of the ensemble, and this difference may be modelled by constructing an ensemble having the same input parameters as previously, but with the variance $\sigma^2$ (where $\sigma^2=(\theta_{FC}-\hat{\theta}_{FC})^2$) as the output parameter. This is known as the 'error ensemble'.

Clearly, the variance will typically be very small, and since the output parameter should ideally be in the range [0, 1], the variance is re-scaled by the algorithm:

$$u_j = 10 \times \sqrt{\text{var}_j} \qquad (6.3a)$$

To recap, the error ensemble has five input parameters, SSC, BD and %OM, and a single output parameter, the re-scaled variance of the original ensemble predictions. Since the inputs are the same as the original ensemble, and the output is closely related to the expected value of $\theta_{FC}$, it is reasonable to assume that the level of complexity of the data to be modelled in the error ensemble is not significantly different from that already modelled in the original ensemble. Thus, the error ensemble is constructed in precisely the same manner as the original ensemble. 20 error ANNs have been trained, each having 25 soil horizons from each texture class,

and the best ten individual error ANNs chosen to construct the error ensemble. This process is repeated ten times, to produce ten error ensembles, designated HCO_Err_Ens_1, HCO_Err_Ens_2, ..., HCO_Err_Ens_10, where HCO denotes that the error ensemble model is a HYPRES Continuous error ensemble, and each of the models are expected to be equivalent.

## 6.4 Error Bar Modelling – Results and Discussion

Comprehensive analyses of various test soil horizons will be conducted in this section and results shown graphically as error bars and 3D error surfaces. These will enable a practical comparison of the two methods of error prediction, and will illustrate their comparative strengths and weaknesses. Additionally, the results of Section 6.4 will be discussed with reference to the results of statistical analyses of the data, detailed in Section 5.2.

### 6.4.1 Bootstrapping Error Bar Results

To visualise errors generated by the Bootstrapping method, I chose the M soil horizon TEST 5 (Table 6.2h) and, varying the BD in increments of 0.01 in the range [0, 4], re-tested ensemble HCO_Ens_10. The standard deviation of the individual ANN predictions were calculated, and the results plotted as $\theta_{FC}$ *versus* BD, Figure 6.4a.

Recall that in HYPRES, the effective values of BD lie in the range [0.76, 2], so HCO_Ens_10 is effectively being asked to extrapolate when making predictions in the ranges [0, 0.76] and [2, 4].

Consider a hypothetical soil horizon, whose BD may be varied to be whatever we wish it to be. When BD = 0, what we would have is a soil horizon containing a single large pore and no soil, and thus for this solid pore, $\theta_{FC} = 1$. Gradually increasing the BD would increase the amount of matter and decrease the amount of pore space, resulting in a concomitant decrease in $\theta_{FC}$. As BD approaches the value of solid mineral (for example, quartz, where BD = 2.6), the pore space approaches zero, and thus $\theta_{FC}$ approaches zero. Beyond the value of BD = 2.6, $\theta_{FC}$ should be equal to zero. In theory then, as BD increases from zero to 2.6, $\theta_{FC}$ decreases from one to zero (continuously, but not necessarily linearly), and is zero for all values of BD > 2.6. For

the real soil horizon of Figure 6.4a, it can be seen that in the lower and upper regions of extrapolation (BD = [0, 0.76] and BD = [2, 4]), the real soil does not, according to the model, behave anything like the hypothetical soil. Regions of low BD should yield high values of $\theta_{FC}$ (close to $\theta_{FC}$ = 1), and high values of BD (i.e. BD > 2.6) should yield values of $\theta_{FC}$ = 0. Since the model produces highly inaccurate predictions in regions of extrapolation, it is very clear that (1) the model should not be asked to extrapolate, or (2) the user of the model should use common sense and exercise extreme caution with such results.

Generally, as BD increases, the ensemble prediction of $\theta_{FC}$ decreases. As BD increases in the range [0, 2], the errors are small and decrease to a minimum at a value of BD = 1.75, indicating that the individual ANNs are in relative agreement as to what $\theta_{FC}$ is for this soil horizon, and confidence is high. Recall from Tables 5.2c and 5.2e, and Figure 6.2i, that the effective range of BD for M soils is [0.76, 1.97], with a mean of 1.5. Thus, it is expected that error bars in this range should be small, indicating that there is a high degree of confidence in the ensemble predictions, and indeed this is the case. As BD increases in the range [2, 4], the error bars diverge rapidly, indicating that the individual ANNs increasingly disagree, and confidence becomes significantly lower, as the extent to which the ANNs are required to extrapolate increases. Also note that for BD > 3, $\theta_{FC}$ < 0.

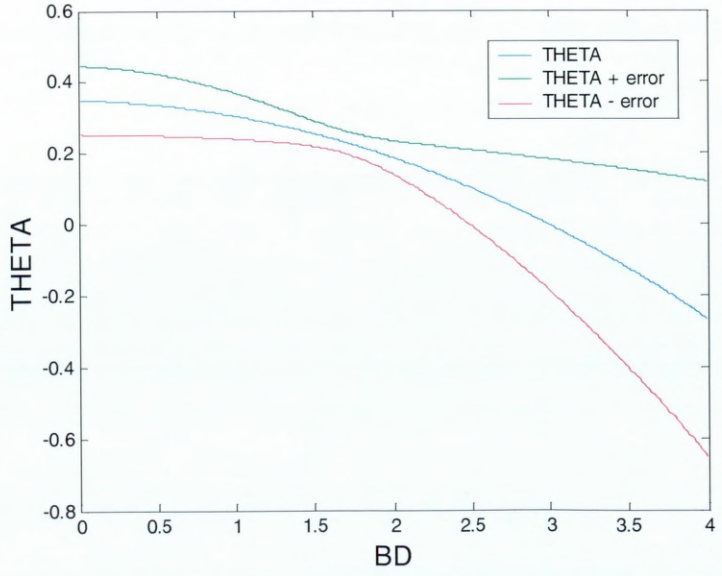



*Figure 6.4a: Effect of varying BD on M soil horizon TEST 5 for ensemble HCO_Ens_10, with Bootstrapping error bars*

This method of visualising has also been applied to the variation of %OM for TEST 5 in increments of 0.1 in the range [0, 15]. Again, HCO_Ens_10 was re-tested for variations in %OM, and the standard deviation of the individual ANN predictions calculated. Results are plotted in Figure 6.4b.

Recall that only soils that have a %OM content in the range [0, 10] were extracted from HYPRES for use in the modelling and testing process, so the ensemble is effectively being asked to extrapolate when making predictions in the range [10, 15].

Generally, as %OM increases, ensemble prediction of $\theta_{FC}$ increases, also the errors on these predictions increase. However, those errors are relatively small in the range [0, 3], increase in the range [3, 10] and increase rapidly in the range [10, 15] as extrapolation occurs. Recall from Tables 5.2c and 5.2e, and Figure 6.2j, that the range of %OM for M soils is [0, 9.4], with a mean of 1.6 (and a median of 1.2), and the majority of measured $\theta_{FC}$ values for M soil horizons fall in the range [0, 3]. Thus, confidence is high in the ensemble predictions in the region [0, 3] where we have the most information about %OM, and confidence decreases significantly as less and less is known about %OM (as %OM is increased from 3 to 15).

Note that for %OM > 13, $\theta_{FC}$ + error > 1.

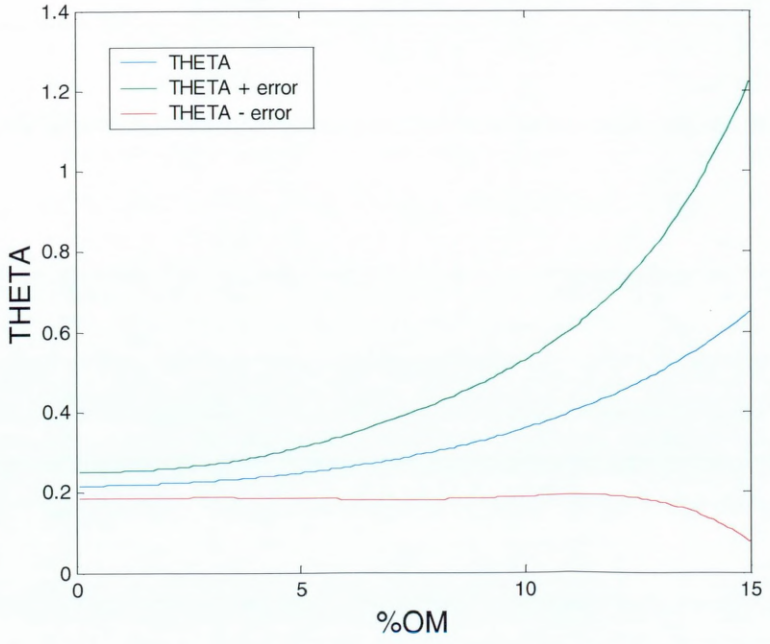



*Figure 6.4b: Effect of varying %OM on M soil horizon TEST 5 for ensemble HCO_Ens_10, with Bootstrapping error bars*

### 6.4.2 ANN Ensemble Error Bar Results

To visualise errors generated by the error ensemble method, I again chose the M soil horizon TEST 5 (see Table 6.2h) and, varying the BD in increments of 0.01 in the range [0, 4], re-tested the ensemble HCO_Ens_10 and an error ensemble chosen at random from the 10 available, HCO_Err_Ens_4. The standard deviation of the individual error ANN predictions were calculated, and the results plotted as $\theta_{FC}$ *versus* BD, Figure 6.4c.

The errors in the range [0, 2] progressively converge, indicating that confidence is increasing, with the exception of a small widening of error bars for BD in the range [1.2, 1.5]. Recall from Tables 5.2c and 5.2e, and Figure 6.2i, that the range of BD for M soils is [0.76, 1.97], with a mean of 1.5. Thus, it is expected that confidence should be high in the range of measured values for BD, and indeed this is the case. However, beyond BD = 2, the errors remain small and stable, with the exception of a marked increase at BD = 2.9. This leads to a high level of confidence in error prediction in the region where the ANNs are extrapolating. This is counter-intuitive.

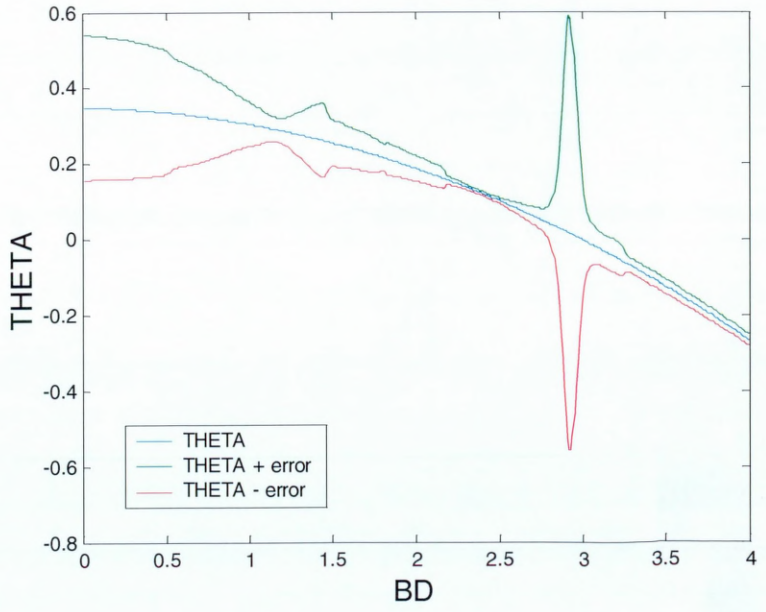



*Figure 6.4c: Effect of varying BD on M soil horizon TEST 5 for ensemble HCO_Ens_10, with error ensemble HCO_Err_Ens_4 error bars*

This method of visualising has also been applied to the variation of %OM for TEST 5 in increments of 0.1 in the range [0, 15]. Again, HCO_Ens_10 and HCO_Err_Ens_4 were re-tested for variations in %OM, and the standard deviation of the individual error ANN predictions calculated. Results are plotted in Figure 6.4d.

There appear to be quite significant anomalous effects on Figure 6.4d. Focussing on the general trend of the error bars as though these effects did not exist, in the range [0, 10], the errors remain small and quite constant, and the error bars widen as the ensemble is required to extrapolate. Recall from Tables 5.2c and 5.2e, and Figure 6.2j, that the range of %OM for M soils is [0, 9.4], with a mean of 1.6 (and a median of 1.2), and the majority of measured $\theta_{FC}$ values for M soil horizons fall within the range [0, 3]. Thus, it is expected that confidence in the ensemble predictions should be high in the range [0, 10], particularly in the range [0, 3], and this is indeed the case. However, there are anomalous widenings of the error bars at %OM = 2.5 and 9 in the interpolation region, with further anomalous features in the region of extrapolation. Although data is scarce when %OM is 9, there is an abundance of data when %OM is 2.5. Thus, the anomalous widenings of the error bars at %OM = 2.5 is a particularly surprising result.

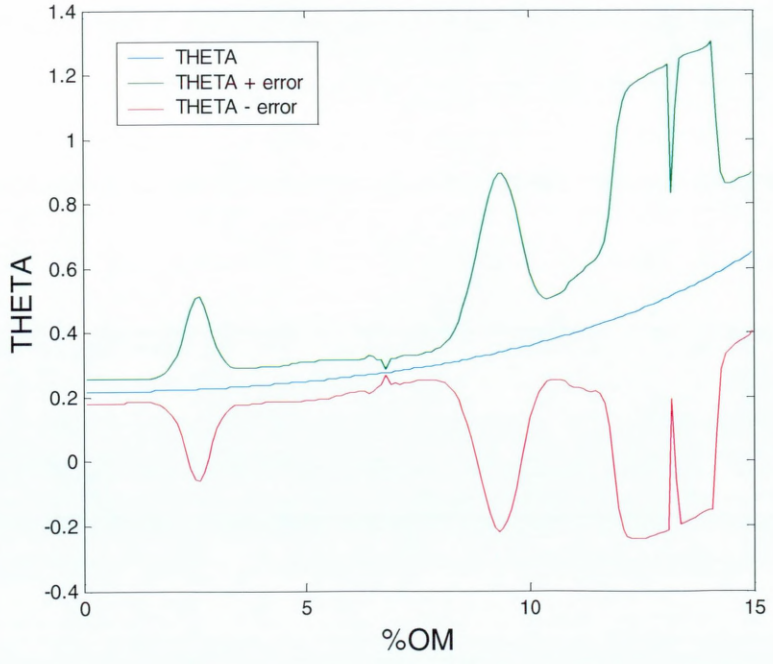



*Figure 6.4d: Effect of varying %OM on M soil horizon TEST 5 for ensemble HCO_Ens_10, with error ensemble HCO_Err_Ens_4 error bars*

Does this mean that this is an effect of the error ensemble method rather than any spurious effects of the data? To answer this question, it is instructive to perform the same tests, but using a different error ensemble. The results of testing HCO_Ens_10 and HCO_Err_Ens_10 (again, chosen at random) with variations in BD are given in Figure 6.4e.

The errors in the range [0, 2] progressively converge in the same manner as with Figure 6.4c, although a little smoother. There is also a small widening in the error bars at BD = 1.6 that is not dissimilar to that seen in Figure 6.4c in the range [1.2, 1.5]. Beyond BD = 2, the errors continue to converge. As with Figure 6.4c, this leads to a high level of confidence in error prediction in the region where the ANNs are extrapolating.

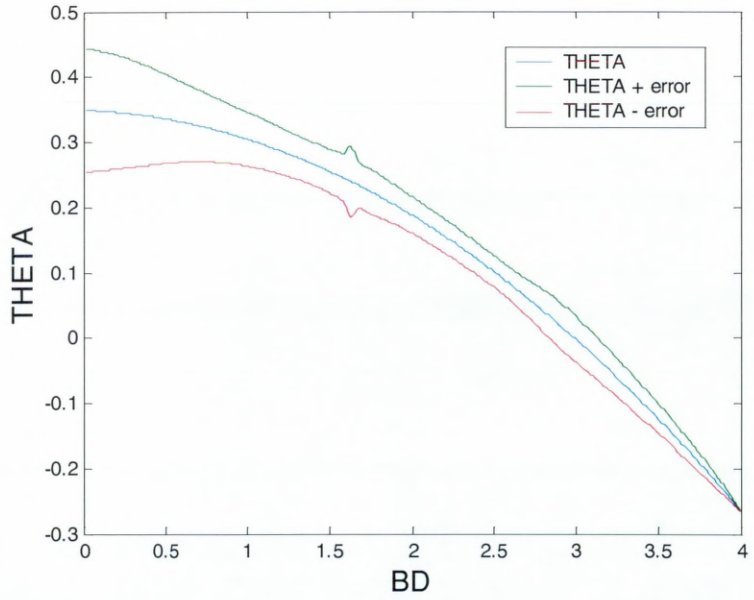



*Figure 6.4e: Effect of varying BD on M soil horizon TEST 5 for ensemble HCO_Ens_10, with error ensemble HCO_Err_Ens_10 error bars*

The results of testing HCO_Ens_10 and HCO_Err_Ens_10 with variations in %OM are given in Figure 6.4f.

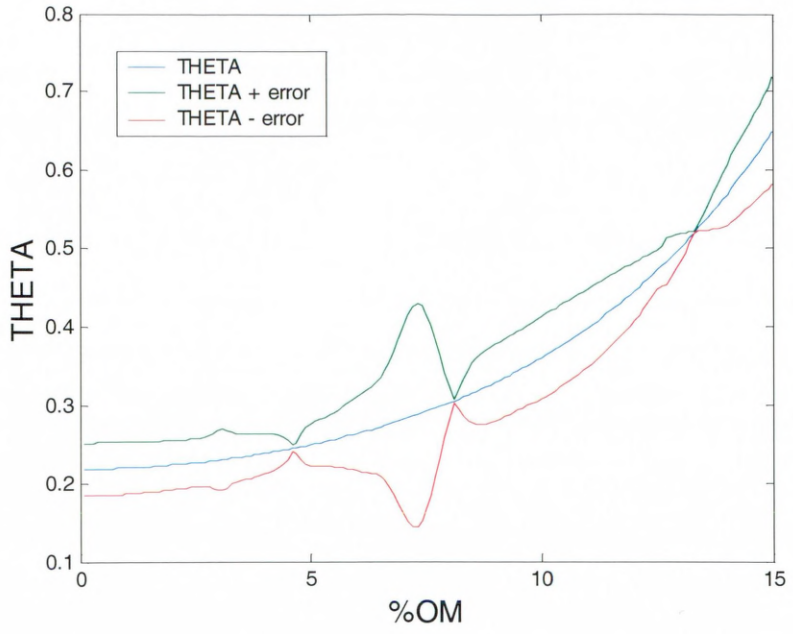



*Figure 6.4f: Effect of varying %OM on M soil horizon TEST 5 for ensemble HCO_Ens_10, with error ensemble HCO_Err_Ens_10 error bars*

Ignoring the anomalous effects, the error predictions remain relatively constant over the entire range tested. With regards to the anomalous features of Figures 6.4d and 6.4f, the widenings do not occur at the same values of %OM, and nor do the regions of pinching.

When inspecting the results of repeat testing using the different error ensembles, the general trend of error prediction in regions of interpolation yields fairly consistent results, although the anomalous results are not consistent and appear quite unhelpful.

### 6.4.3 Error Bar Results Comparison and Discussion

Re-plotting the results of Figures 6.4a and 6.4c (Figure 6.4g), and 6.4a and 6.4e (Figure 6.4h), allow the direct comparison of the error ensemble and Bootstrapping methods of error prediction for BD for the two error ensembles HCO_Err_Ens_4 and HCO_Err_Ens_10 tested, respectively. Since we are only interested in regions of interpolation, these results have been plotted in the BD range [0, 2].

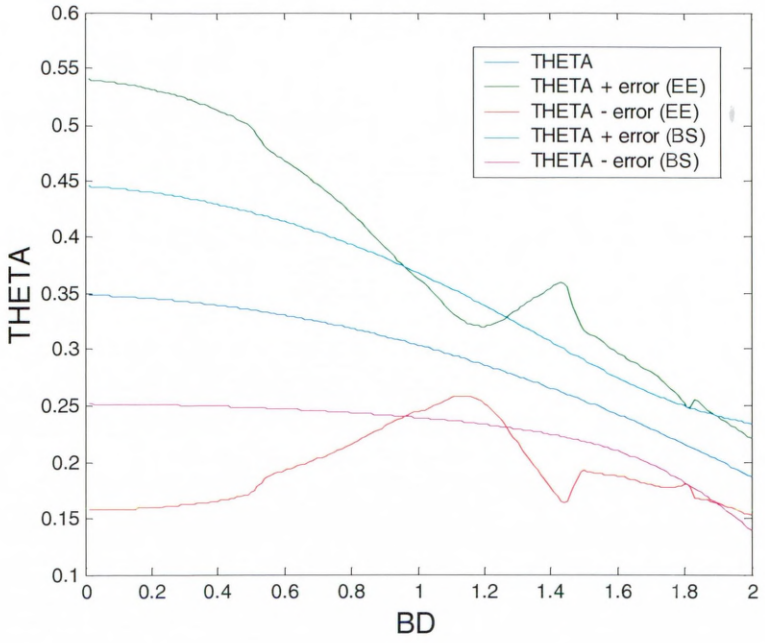



*Figure 6.4g: Effect of varying BD on M soil horizon TEST 5 for ensemble HCO_Ens_10, with error ensemble HCO_Err_Ens_4 (EE) and Bootstrapping (BS) error bars*

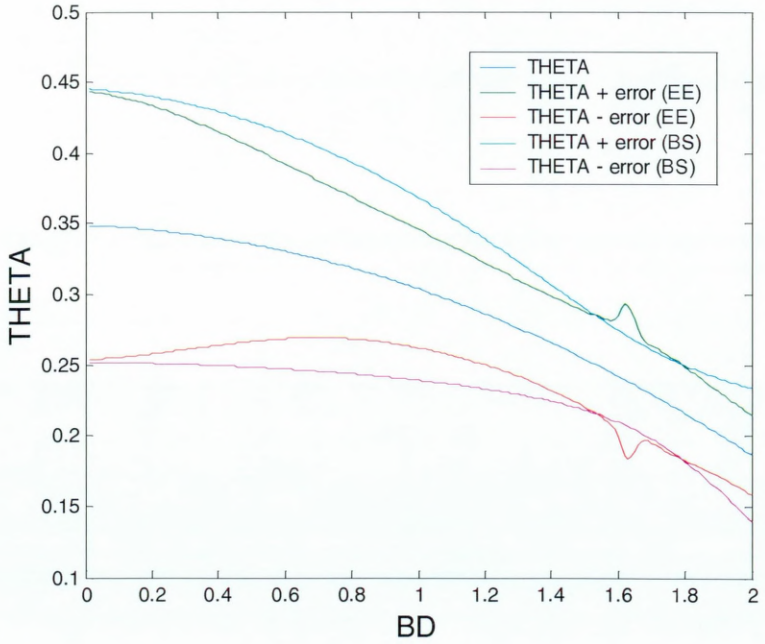



*Figure 6.4h: Effect of varying BD on M soil horizon TEST 5 for ensemble HCO_Ens_10, with error ensemble HCO_Err_Ens_10 (EE) and Bootstrapping (BS) error bars*

Both methods of error prediction provide a decrease in errors as BD increases, for the error ensemble method, regardless of the error ensemble that was used, and generally, the error ensemble method makes larger predictions of errors than does the Bootstrapping method, for this particular M soil horizon. The Bootstrapping method describes a very smooth error function, whereas the error ensemble method seems to be rather inconsistent, making quite spurious predictions at seemingly random sections of the dataspace.

Re-plotting the results of Figures 6.4b and 6.4d (Figure 6.4i), and 6.4b and 6.4f (Figure 6.4j), allow the direct comparison of the error ensemble and Bootstrapping methods of error prediction for %OM for the two error ensembles HCO_Err_Ens_4 and HCO_Err_Ens_10 tested, respectively. Since we are only interested in regions of interpolation, these results have been plotted in the %OM range [0, 10].

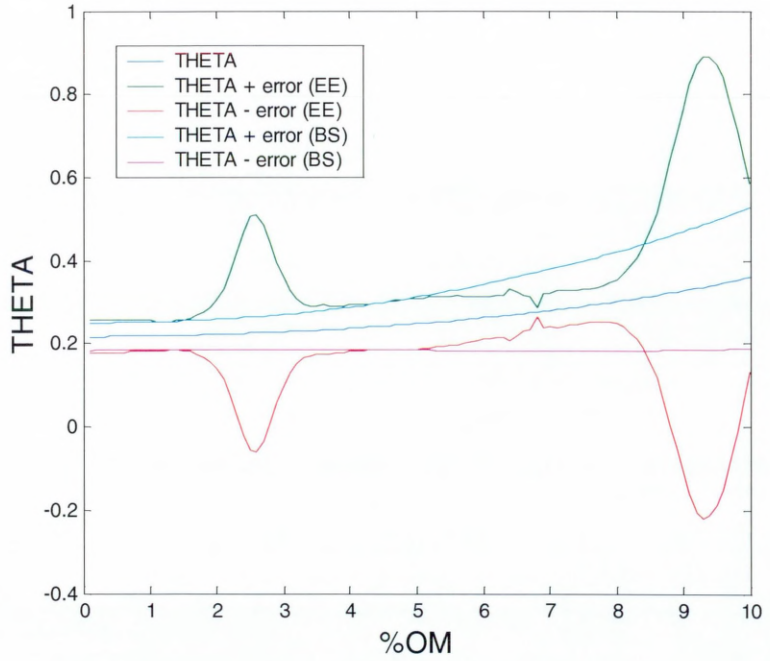



*Figure 6.4i: Effect of varying %OM on M soil horizon TEST 5 for ensemble HCO_Ens_10, with error ensemble HCO_Err_Ens_4 (EE) and Bootstrapping (BS) error bars*

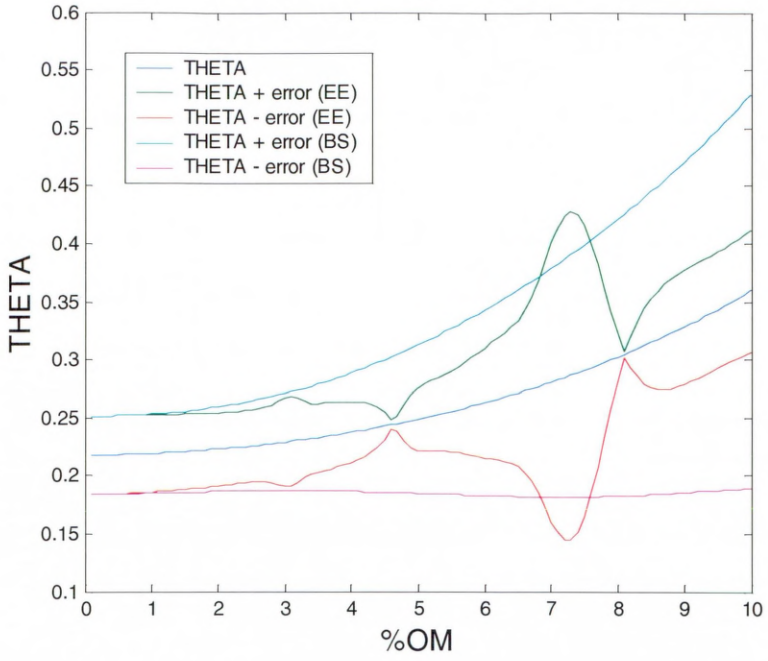



*Figure 6.4j: Effect of varying %OM on M soil horizon TEST 5 for ensemble HCO_Ens_10, with error ensemble HCO_Err_Ens_10 (EE) and Bootstrapping (BS) error bars*

Plotting $\theta_{FC}$ *versus* BD and %OM as a 3D surface plot allows the reader to see how $\theta_{FC}$ varies for both of these parameters. To do this, SSC of TEST 5 are held constant, whilst both BD and %OM are varied within their respective training ranges for model HCO_Ens_10, in increments of 0.02 and 0.1, respectively. The results are shown in the 3D surface plot of Figure 6.4k.

These results show that, for this particular test soil sample, $\theta_{FC}$ is a smoothly decreasing and increasing function of BD and %OM, respectively. The ensemble model HCO_Ens_10 does not make any anomalous predictions of $\theta_{FC}$ on varying BD and %OM for this soil horizon and, since none are expected, this is a good sign. However, caution should be recommended when reading the graph of Figure 6.4k, since soil samples that have zero BD correspond to a solid pore composed entirely of air. Thus, the model has little pedologic meaning for soils having BD < 0.76 (since the model would be extrapolating), and is of no use whatsoever when BD $\approx$0.

It is interesting to note that, according to Table 5.2e, on the average, as %OM increases, $\theta_{FC}$ increases, and also that, on the average, as BD increases, $\theta_{FC}$

decreases, and this is exactly what we see in Figure 6.4k. In addition, the CV of %OM is high, Table 5.2g, indicating that %OM is highly variable, and perhaps more difficult to model accurately than parameters that are more consistent.

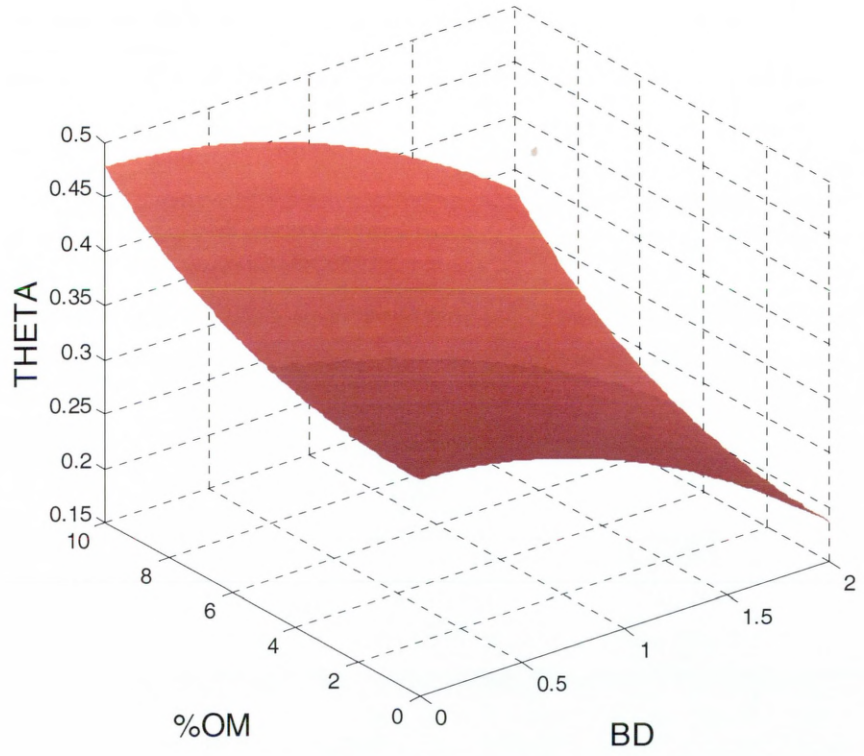



*Figure 6.4k: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble model HCO_Ens_10*

For the same soil horizon, the error ensemble HCO_Err_Ens_4 may be tested in the same manner to determine which regions of this prediction surface are predicted with confidence (low error), and which areas should be treated with caution. The results of this are shown in the 3D surface plot of Figure 6.4l, which plots error *versus* BD and %OM.

As expected from earlier 2D plots of error *versus* BD and error *versus* %OM, there are a number of anomalous features on this graph. However, the error surface is relatively flat and featureless for most values of BD and %OM, and errors only become large when %OM > 8, and when BD > 1.5 and %OM < 4. Recall that the effective ranges of BD and %OM are [0.76, 1.97] and [0, 3], respectively.

Considering Table 5.2f, we see that, since %OM is strongly right-skewed, most of the %OM data is for low-valued %OM (in the range [0, 3]), and hence predictions for high-valued %OM (in the range [3, 10]) should display reduced confidence. This reflects the fact that the ensemble has comparatively little knowledge of these particular regions of the dataspace, and is extrapolating between distant data points, and this is what we see in Figure 6.4l. Similarly, since BD is left-skewed, predictions for high-valued BD should be high in confidence, and in general, this is the case, although there is a large anomalous effect on Figure 6.4l for high-valued BD (at low %OM).

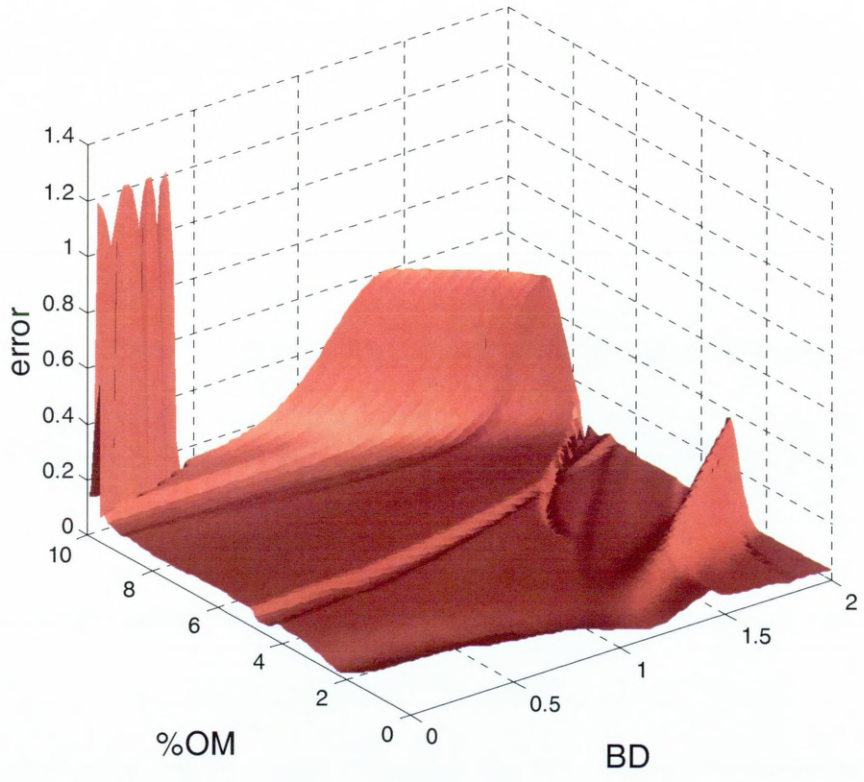



*Figure 6.4l: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble HCO_Ens_10 and error ensemble HCO_Err_Ens_4*

Figure 6.4m shows the same tests as those of Figure 6.4l, but using error ensemble HCO_Err_Ens_10.

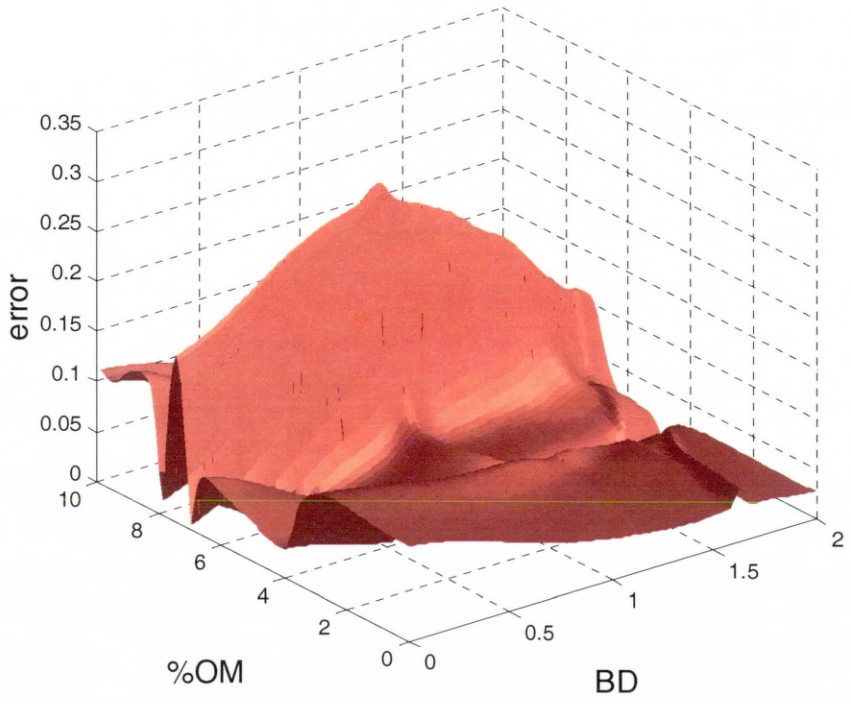



*Figure 6.4m: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble HCO_Ens_10 and error ensemble HCO_Err_Ens_10*

As with Figure 6.4l, errors on predictions become larger as %OM increases, although the magnitude of those errors differ considerably.

While Figure 6.4m may appear to differ significantly with Figure 6.4l, the same general features are in evidence. High values of %OM produce the largest errors, and high values of BD produce the smallest errors, as predicted by Table 5.2f.

In order to assess how well this error surface matches with the error surface that would be predicted by the Bootstrapping method, the same procedure has been carried out on TEST 5 and the results plotted as Figure 6.4n.

The Bootstrapping method of error prediction appears to show a smooth relationship between prediction error, BD and %OM. Note that the errors are greatest when %OM is large. This is the same result as in the error ensemble method, for all error ensembles trained, albeit that the magnitude of the predicted errors differ considerably. Note also that the region of smallest error is when BD is large and %OM is small, corresponding to the effective ranges for BD and %OM, [0.76, 1.97] and [0, 3], respectively. This region produced similar general error predictions in the error ensemble method, although some anomalous effects were in evidence.

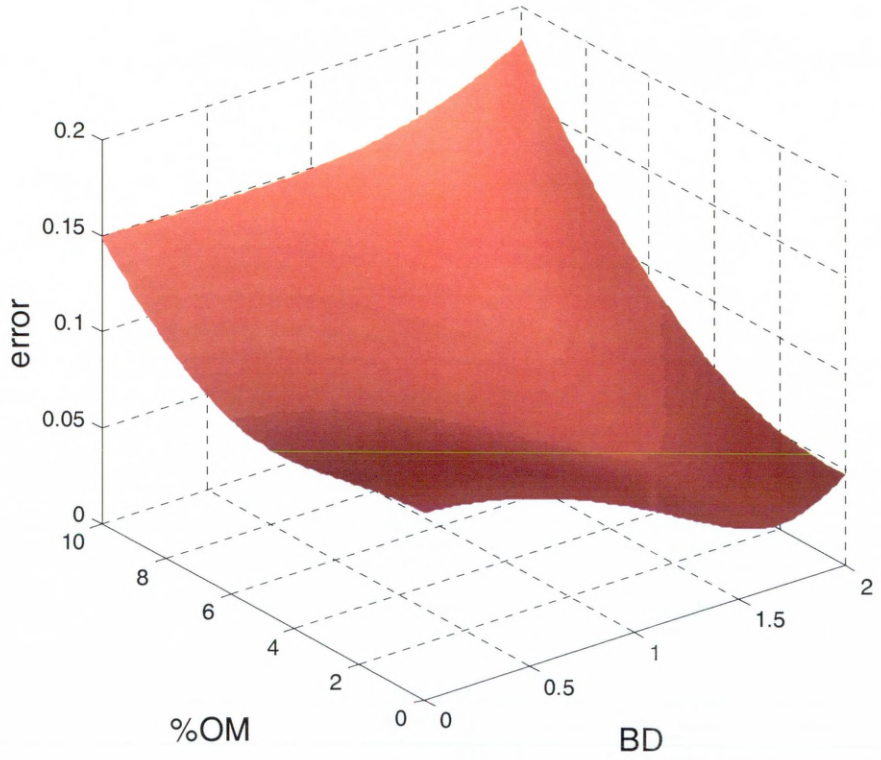



*Figure 6.4n: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble HCO_Ens_10 and the Bootstrapping method of error prediction*

Overall, both methods of error prediction provide a decrease in errors as BD increases, regardless of the error ensemble that was used, although the error ensemble method is not consistent as to whether the predictions of errors are smaller or larger than the Bootstrapping method, for this particular M soil horizon. The Bootstrapping method describes a very smooth error function, whereas the error ensemble method gives rise to many anomalous features at various regions of the dataspace.

In theory, the error ensemble method appears more favourable than does the Bootstrapping method, since it is making a prediction of the distance from the ensemble prediction of $\theta_{FC}$ to the expected value of $\theta_{FC}$, and not simply the spread of predictions of the individual ANNs. However, in practice, unlike the Bootstrapping method, the error ensemble method does not give smooth, consistent predictions.

Despite the difficulties I have found here with the error ensemble method, Roberts *et al.* (1996), Penny and Roberts (1997) and Penny and Roberts (1998) have used this method of prediction to great effect.

## Chapter Conclusions

Typically with ANN-PTFs, the practitioner will split the available data into three sub-sections; one for training, one for validation and one for testing. Each sub-section will retain complete independence from the others, and is often extracted from the available data randomly. In the example where there are 1200 available soil horizons, and the dataset is split in the ratio 1:1 to provide the test set, and the remainder is split in the ratio 2:1 to provide the training and validation sets, respectively, this yields the training:validation:test ratio of 400:200:600. At this point an ANN will be trained and tested with these data. Considering that of the HYPRES data chosen for modelling (Table 5.1b) only 4% belongs to the VF texture class, whilst 34% belongs to the M texture class, this means that typically 16 VF and 136 M soil horizons will be used for training, whilst 24 VF and 204 M soil horizons will be used for testing in the above example. This method raises many critical questions, such as 'how many soil horizons are sufficient for effective training?', and 'how many soil horizons are required for stringent testing?'. Also, both the training and test sets are clearly biased. I have demonstrated in this thesis that by using the ensemble method of ANN-PTF construction we may reduce overall prediction errors compared to the single ANN method by *using* the bias/variance trade-off, rather than being *restricted* by it. This involves choosing data with equal frequency from each sub-grouping of the dataset, in this case by texture class, to increase variance and decrease bias in the individual ANN models. On combining the individual ANNs to produce the ensemble, the variance is eliminated, thereby decreasing prediction errors. I have shown that ten ANN ensemble members, each trained with 25 soil horizons Bootstrapped from each texture class produces results that are equivalent to a single ANN trained with three times the amount of data. I have also shown that the number of test horizons required for stringent testing to two decimal points is not less than 150 for each soil texture class, and that overall the ensemble model is stable and unbiased.

On performing sensitivity analyses on the ensemble, it was demonstrated that a continuous ANN-PTF ensemble of water retention at field capacity is insensitive to

the loss of any individual input parameter, with the exception of BD. Omitting BD from the ensemble model results in a decrease in predictive ability, and this ability is weakened further when %OM is also omitted from the model. Models that have no soil textural component (i.e. only BD and %OM are present) have significantly decreased predictive abilities.

In this chapter, two possible methods of producing error bars on individual predictions were investigated; the Bootstrapping method, and the error ensemble method. The Bootstrapping method involves using the calculation of the spread of the predictions of individual ANNs in the ensemble to express a level of confidence (measure of agreement between individual members) about the ensemble prediction. The error ensemble prediction involves calculating the variance between the expected and ensemble-predicted values of $\theta_{FC}$, and using this as the output parameter in an error ensemble that is constructed in precisely the same way as the continuous ANN-PTF ensemble. Tests show that for the HYPRES continuous ANN-PTF ensemble of $\theta_{FC}$, the Bootstrapping and error ensemble methods produce similarly confident results in the region of high BD and low %OM, as expected, albeit that the error ensemble method did produce some anomalous effects.

# Chapter 7

# Class PTF of Soil Water Retention – an ANN Ensemble Method

The previous chapter demonstrated the procedures that are required in order to optimise a continuous ANN-PTF ensemble of water retention at field capacity. Using the same procedures, this chapter documents the construction of an optimised class ANN-PTF ensemble of water retention at field capacity. Sensitivity analyses are conducted on the ensemble to determine which of the input parameters are of most importance to such a model. In addition, the two methods of producing error bars on individual predictions detailed in Chapter 6 are employed here, and the results of these two competing methods will be discussed in conjunction with the results of statistical analyses conducted in Section 5.2.

## 7.1 Preliminary Investigations and Optimisation Procedures

Section 6.1 detailed the various parameters that need to be optimised in order to build an ANN-PTF by the ensemble method that generalises well for unseen data, and maximises information usage, whilst minimising the amount of data that is required. These procedures have been used to build a continuous ANN-PTF ensemble of water retention at field capacity (Section 6.2), and now these same procedures are used to construct a class ANN-PTF ensemble of water retention at field capacity. Since M textured soils are most abundant in the data selected for modelling (Table 5.1b), this class is used to build the class PTF. Modelling of class PTFs of the other texture classes will be discussed further in Chapter 11.

### 7.1.1 Optimum Number of ANN Ensemble Members

Having established in Section 6.2.2 that the optimum number of ANN ensemble members and optimum numbers of soil horizons per ANN may be determined independently from each other, the optimum number of members to employ in the ensemble has been determined following the protocol set out in Section 6.1.4. The

plot of Figure 7.1a shows the result of this investigation, with RMSE plotted against the number of ANN ensemble members.

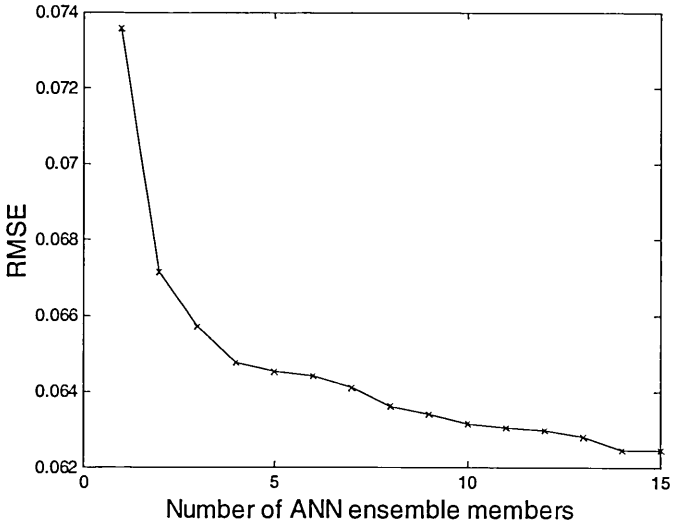



*Figure 7.1a: Graph of RMSE* versus *number of ANN ensemble members*

Since the number of ensemble members used in constructing the continuous PTF of Section 6.2 was ten, it seems reasonable to assess if this number should be used in constructing the class PTF. Calculating from Figure 7.1a, the RI (see Equation 6.1c) of ten members over one is 14%, whilst the RI of 15 members over one is 15%. The evidence from Figure 7.1a and the RIs calculated, show that using ten ensemble members is adequate.

### 7.1.2 Evidence for the Optimum Number of Soil Horizons per ANN

The optimum number of M soil horizons to use per ANN ensemble member has been determined following the protocol set out in Section 6.1.7. This procedure has been repeated ten times with the RMSE results averaged to provide smoothing. The plot of Figure 7.1b shows the results of this investigation, with RMSE plotted against the number of soil horizons per ANN ensemble member. Additionally, the same procedure has been followed to determine the plot of ME against the number of soil horizons per ANN ensemble member (Figure 7.1c). Only the overall results are shown on Figures 7.1b and 7.1c, not the accuracy or reliability results.

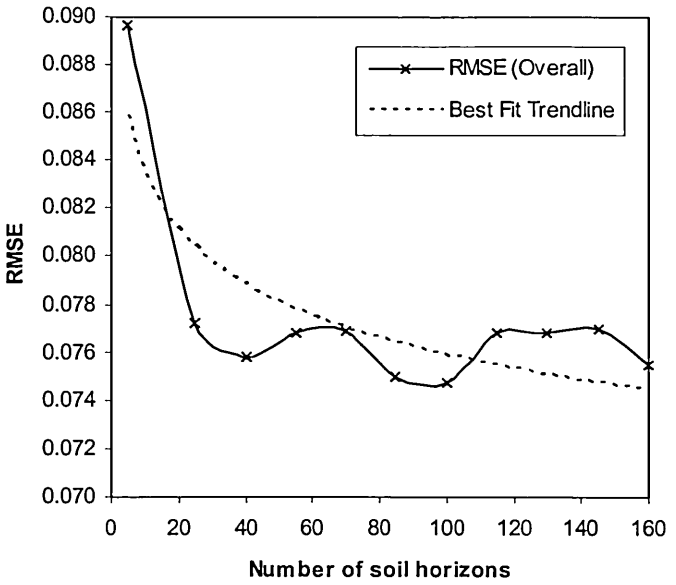



*Figure 7.1b: Results of tests to obtain evidence for the optimum number of M soil horizons per ANN*

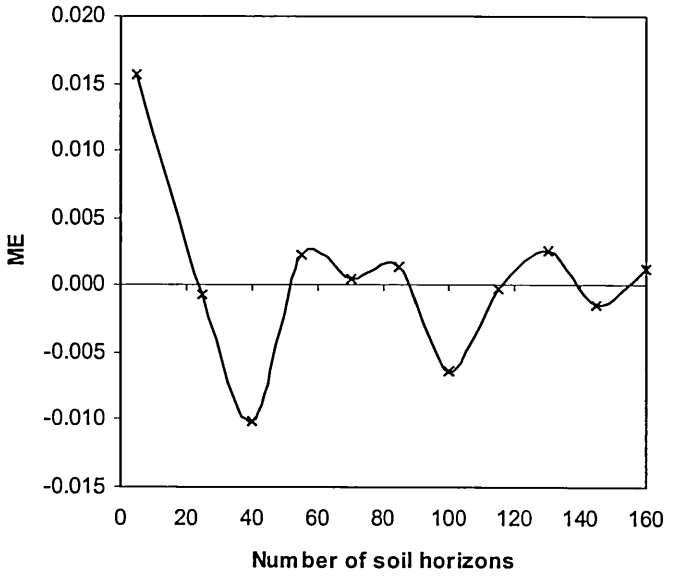



*Figure 7.1c: Results of tests to obtain evidence for the optimum number of M soil horizons per ANN*

The best fit trendline in Figure 7.1b continues to decrease as more soil horizons are added, but the rate of decrease slows down beyond about 80 soil horizons. Considering the best fit trendline, the results seem to suggest that only small improvements may be made with greater than about 80 or 100 soil horizons.

Quantitatively, the relative improvement of 100 horizons over 5 horizons is 11.6%, whereas the relative improvement of 160 horizons over 5 horizons is 12.8%. Figure 7.1c suggests that when more than 50 or 60 soil horizons are used, the model bias is very small, being less than 10% of the corresponding RMSE values.

The number of M soil horizons that will be used in the class ANN-PTF ensemble of water retention at field capacity is 100.

## 7.2 Class PTF – Results and Discussion

Section 6.2.2 detailed the protocol for constructing a continuous ANN-PTF ensemble of soil water retention at field capacity, and now these same procedures are used to construct a class ANN-PTF ensemble of water retention at field capacity. The only difference in the procedures is that in the continous PTF, 25 soil horizons per ANN member were used from each texture class, whilst in the class PTF, 100 soil horizons from the M soil texture class are used.

### 7.2.1 ANN Ensemble Soil Water Retention Class PTF Results

Table 7.2a shows the RMSE and ME results (all tests) of each of the ten ensembles, denoted HCL_Ens_1, HCL_Ens_2, ..., HCL_Ens_10, where HCL denotes that the model is a HYPRES Class PTF.

| | M | |
|---|---|---|
| | **RMSE** | **ME** |
| **HCL_Ens_1** | 0.061 | 0.0010 |
| **HCL_Ens_2** | 0.065 | -0.0040 |
| **HCL_Ens_3** | 0.059 | 0.0033 |
| **HCL_Ens_4** | 0.060 | 0.0038 |
| **HCL_Ens_5** | 0.065 | 0.0020 |
| **HCL_Ens_6** | 0.061 | 0.0015 |
| **HCL_Ens_7** | 0.060 | 0.0013 |
| **HCL_Ens_8** | 0.060 | -0.0010 |
| **HCL_Ens_9** | 0.062 | -0.0011 |
| **HCL_Ens_10** | 0.061 | -0.0008 |

*Table 7.2a: RMSEs and MEs of 10 ensembles*

In Section 6.1.3 it was established that, for the amount of data that is available for these investigations, the results will probably only be reliable to two decimal places, however, quoting to three decimal places demonstrates the stability of the ensembles, each varying little (RMSE) in terms of their predictive abilities.

Here, all MEs are at least one order of magnitude smaller than the corresponding RMSE results, indicating that the bias in each of the models is very small. Also note that the sign of the ME is not consistent, indicating that $\theta_{FC}$ is neither consistently under- nor over-estimated.

Table 7.2b shows the average results, both RMSE and ME, of the ten ensembles.

| | **M** |
|---|---|
| **Number of tests** | 952 |
| **RMSE** | 0.061 |
| **ME** | 0.0006 |

*Table 7.2b: RMSE and ME results, averaged over 10 ensemble models*

Figure 7.2a shows how the ensemble-predicted values correspond with the expected values of $\theta_{FC}$, for HYPRES class ANN-PTF ensemble model HCL_Ens_3 of the M texture class.

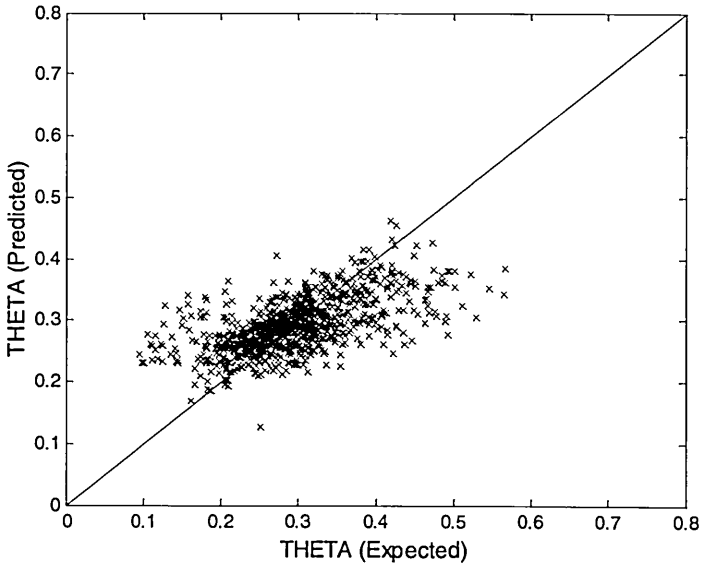



*Figure 7.2a: Comparison of expected and ensemble-predicted $\theta_{FC}$ for ensemble model HCL_Ens_3 of the M texture class*

There is quite a good correspondence between ensemble-predicted and expected values of $\theta_{FC}$ for M textured soil horizons, with most values falling on or near to the expected 1:1 correspondence.

A comparison of the continuous and class ensemble-predicted and expected values of $\theta_{FC}$ for the HYPRES M texture class will be given in Section 7.2.3.

To visualise the effect of varying BD on predictions of $\theta_{FC}$ for the HYPRES class ANN-PTF ensemble, the same five horizons as those of Table 6.2h were used to re-test HCL_Ens_3. BD was varied in increments of 0.01 in the range [0, 2], HCL_Ens_3 was re-tested, and the results plotted as $\theta_{FC}$ *versus* BD. The soil horizons that were chosen are shown in Table 6.2h. The results of such tests are shown graphically in Figure 7.2b, additionally, the BD has been plotted against measured values of $\theta_{FC}$ for all the HYPRES M soil horizons used in the modelling process, and a linear regression line plotted.

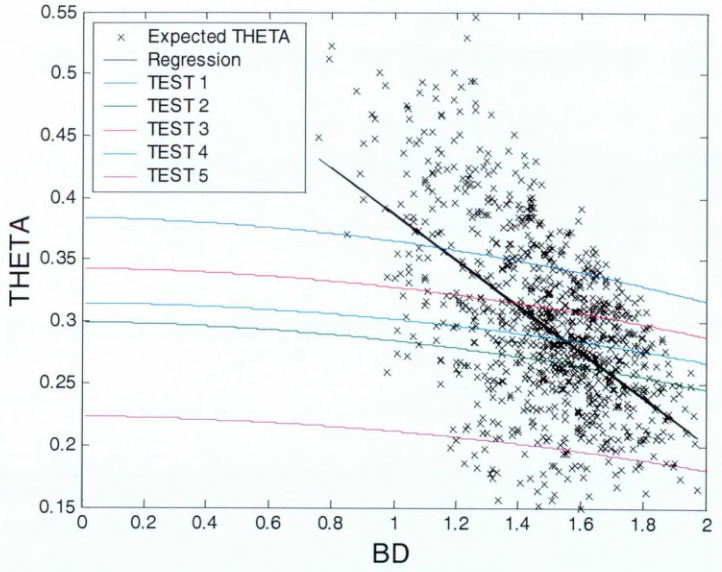



*Figure 7.2b: Effect of varying BD on five randomly chosen M soil horizons for ensemble HCL_Ens_3, with measured values of $\theta_{FC}$ and regression line*

As BD increases, $\theta_{FC}$ decreases for all M soils tested, and the magnitude of the decrease is small and consistent. From Tables 5.2c and 5.2e, the range of BD for M soils is [0.76, 1.97], with a mean of 1.5, and the range and mean of $\theta_{FC}$ for M soils

are [0.1, 0.57] and 0.3, respectively. The regression line on Figure 7.2b shows that one would expect that as the BD is increased $\theta_{FC}$ should decrease, and this is exactly what is seen. It should be noted that the model has no pedologic meaning for soils having BD < 0.76 (since the model would be extrapolating). Additionally, the gradient of the regression line in Figure 7.2b is quite different from the gradients of the five test samples shown in the range [0.8, 2].

This procedure has also been repeated with respect to %OM, and incremented by 0.1 in the range [0, 10]. The results of $\theta_{FC}$ *versus* %OM are shown in Figure 7.2c, additionally, the %OM has been plotted against measured values of $\theta_{FC}$ for all the M soils used in the modelling process, and a linear regression line plotted.

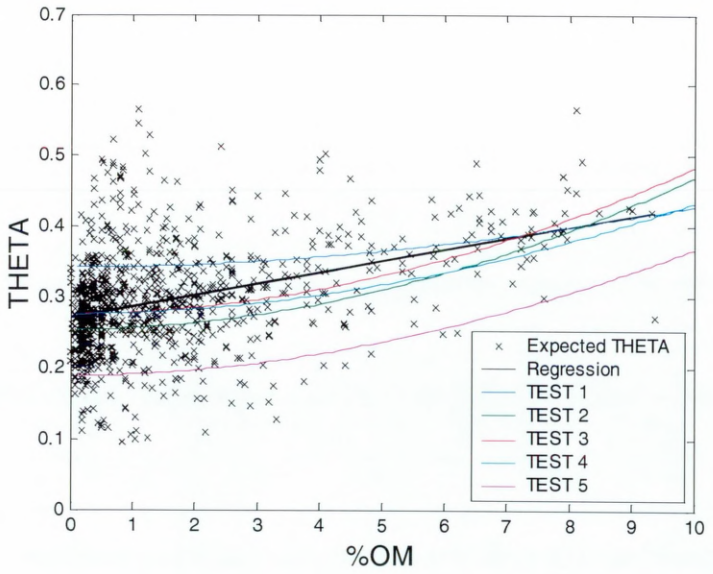



*Figure 7.2c: Effect of varying %OM on five randomly chosen M soil horizons for ensemble HCL_Ens_3, with measured values of $\theta_{FC}$ and regression line*

As %OM is increased, $\theta_{FC}$ increases for all M soils tested, although the magnitude of the effect differs from horizon to horizon.

From Tables 5.2c and 5.2e, the range of %OM for M soils is [0, 9.4], with a mean of 1.6 (and a median of 1.2), and the range of $\theta_{FC}$ for M soils is [0.1, 0.57], with a mean of 0.3. The regression line on Figure 7.2c shows that one would expect that as the %OM is increased $\theta_{FC}$ should increase, and this is exactly what is seen for all sample tests. It should be noted that the majority of %OM values for M soil horizons

at $\theta_{FC}$ fall in the range [0, 3], and since there is much less data in the range [3, 10], it should be expected that the model produces more reliable predictions for %OM < 3. The gradient of the regression line in Figure 7.2c is not significantly different from the gradients of the five test samples shown.

Overall, for the class PTF tested with selected M soil horizons, it appears that varying BD has little effect on the prediction of $\theta_{FC}$, whereas varying %OM has a significant effect, increasing the prediction of $\theta_{FC}$ by as much as 100% (TEST 5 at %OM = 10, compared with %OM = 0).

### 7.2.2 Sensitivity Analyses of the Ensemble

Sensitivity analyses have been conducted on the class PTF in the same manner as in Section 6.2.5. Each of the five input parameters has been omitted in turn from the class PTF ensemble model, also SSC was omitted, and BD & %OM, and then each of these re-trained and re-tested in precisely the same way as in the ensemble method. This procedure is repeated 10 times, and the results averaged. The RMSE results of testing are shown in Table 7.2c. The results are also shown graphically in Figure 7.2d. Additionally, the RMSE results of the ensemble with no parameter omissions are repeated for comparative purposes (results averaged over 10 procedural repeats). The ME results varied very little from model to model and are not shown here.

| Input parameters used | | | | | RMSE |
|---|---|---|---|---|---|
| % Cl | % Si | % Sa | BD | %OM | M |
| ✓ | ✓ | ✓ | ✓ | ✓ | 0.061 |
| | ✓ | ✓ | ✓ | ✓ | 0.060 |
| ✓ | | ✓ | ✓ | ✓ | 0.060 |
| ✓ | ✓ | | ✓ | ✓ | 0.059 |
| ✓ | ✓ | ✓ | | ✓ | 0.062 |
| ✓ | ✓ | ✓ | ✓ | | 0.061 |
| ✓ | ✓ | ✓ | | | 0.068 |
| | | | ✓ | ✓ | 0.071 |

*Table 7.2c: RMSE results of sensitivity analyses of the ensemble method, averaged over 10 models*

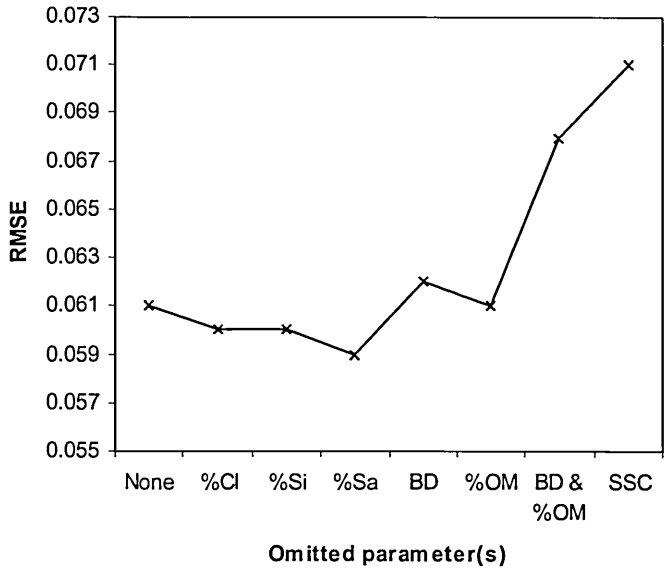



*Figure 7.2d: RMSEs of ensemble sensitivity analyses, averaged over 10 models*

Omitting an individual parameter has no significant effect on the class PTF as a whole, however, omitting BD & %OM, the predictions deteriorate, and the deterioration increases with the omission of SSC.

## 7.3 Error Bar Modelling – Results and Discussion

Analyses of the class PTF, error ensemble and Bootstrapping methods will be conducted in this section using M soil horizon TEST 5 (Table 6.2h), and results shown graphically as error bars and 3D error surfaces. These will enable a practical comparison of the two methods of error prediction, and will illustrate their comparative strengths and weaknesses. Additionally, the results of Section 7.2.1 will be discussed with reference to the results of statistical analyses of the data, detailed in Section 5.2.

To visualise errors generated by both the Bootstrapping and error ensemble methods, the BD of the M soil horizon TEST 5 was varied in increments of 0.01 in the range [0, 2], and the ensemble HCL_Ens_3 and error ensemble HCL_Err_Ens_1 re-tested. The standard deviation of the individual ANN predictions and the error ensemble predictions were calculated, and the results plotted as $\theta_{FC}$ *versus* BD,

Figure 7.3a, allowing the direct comparison of the error ensemble and Bootstrapping methods of error prediction.

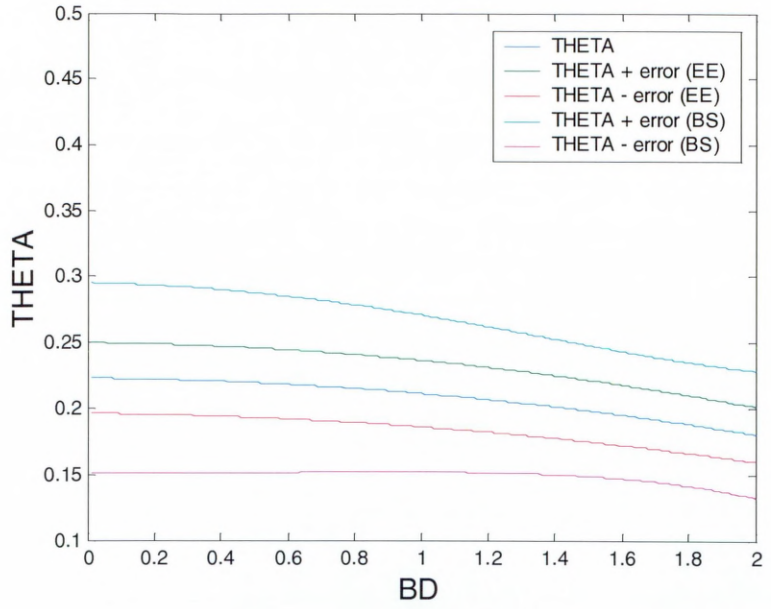



*Figure 7.3a: Effect of varying BD on M soil horizon TEST 5 on ensemble HCL_Ens_3, with error ensemble HCL_Err_Ens_1 (EE) and Bootstrapping (BS) error bars*

Generally, as BD increases, the ensemble prediction of $\theta_{FC}$ decreases. As BD increases in the range [0, 2], the error ensemble errors are small and constant, indicating that the individual ANNs are in relative agreement as to what $\theta_{FC}$ is for this soil horizon, and confidence is high. Recall from Tables 5.2c and 5.2e, and Figure 6.2i, that the effective range of BD for M soils is [0.76, 1.97], with a mean of 1.5. Thus, it is expected that error bars in this range should be small, indicating that there is a high degree of confidence in the ensemble predictions, and indeed this is the case. Indeed, the error bars are constant throughout the entire range tested, indicating that confidence is high for all predictions of $\theta_{FC}$ when BD is in the range [0, 2]. The Bootstrapping method produces error bars that are similar with regard to their constancy to the error ensemble method, albeit that the magnitude of the Bootstrapping error is two to three times larger than the corresponding error ensemble prediction.

This method of visualising has also been applied to the variation of %OM for TEST 5 in increments of 0.1 in the range [0, 10]. Again, HCL_Ens_3 and HCL_Err_Ens_1 were re-tested for variations in %OM, and the standard deviation of the individual ANN predictions and error ensemble predictions calculated. Results are plotted in Figure 7.3b.

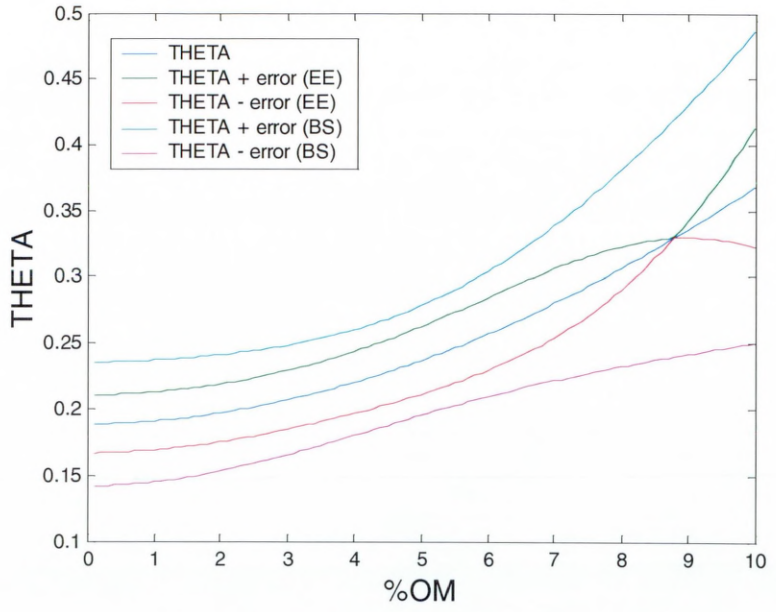



*Figure 7.3b: Effect of varying %OM on M soil horizon TEST 5 on ensemble HCL_Ens_3, with error ensemble HCL_Err_Ens_1 (EE) and Bootstrapping (BS) error bars*

Generally, as %OM increases, ensemble prediction of $\theta_{FC}$ increases. The error ensemble method of error prediction produces error bars that are small and constant, except for a constriction at %OM = 8.5. Recall from Tables 5.2c and 5.2e, and Figure 6.2j, that the range of %OM for M soils is [0, 9.4], with a mean of 1.6 (and a median of 1.2), and the majority of measured $\theta_{FC}$ values for M soil horizons fall in the range [0, 3]. Thus, it is expected that there should be high confidence in the ensemble predictions in the region [0, 3] where we have the most information about %OM. Narrow error ensemble error bars in this region confirm that confidence is high. Indeed, confidence is high throughout the entire tested range of %OM values for the error ensemble method. The Bootstrapping method produces error bars that are wider than their corresponding error ensemble predictions, and diverge when %OM > 5.

Plotting $\theta_{FC}$ *versus* BD and %OM as a 3D surface plot allows the reader to see how $\theta_{FC}$ varies for both of these parameters. To do this, SSC of TEST 5 are held constant, whilst both BD and %OM are varied within their respective training ranges for model HCL_Ens_3, in increments of 0.02 and 0.1, respectively. The results are shown in the 3D surface plot of Figure 7.3c.

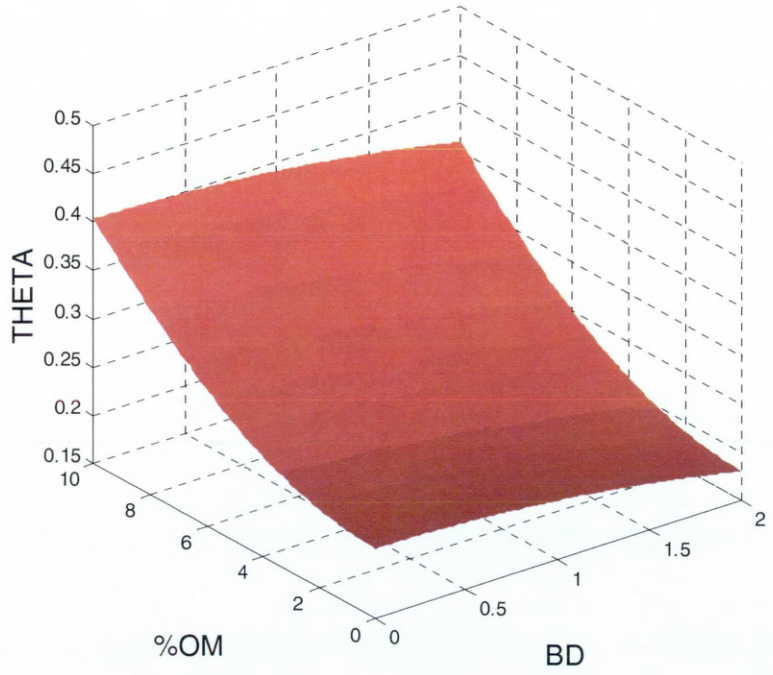



*Figure 7.3c: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble HCL_Ens_3*

These results show that, for this particular test soil sample, $\theta_{FC}$ is a smoothly decreasing and increasing function of BD and %OM, respectively. The ensemble model HCL_Ens_3 does not make any anomalous predictions of $\theta_{FC}$ on varying BD and %OM for this soil horizon and, since none are expected, this is a good sign. However, caution should be recommended when reading the graph of Figure 7.3c, since soil samples that have zero BD correspond to a solid pore composed entirely of air. Thus, the model has little pedologic meaning for soils having BD < 0.76 (since the model would be extrapolating), and is of no use whatsoever when BD is zero.

It is interesting to note that, according to Table 5.2e, on the average, as %OM increases, $\theta_{FC}$ increases, and also that, on the average, as BD increases, $\theta_{FC}$ decreases, and this is exactly what we see in Figure 7.3c. In addition, the CV of %OM is high, Table 5.2g, indicating that %OM is highly variable, and perhaps more difficult to model accurately than parameters that are more consistent.

For the same M soil horizon, TEST 5, the error ensemble HCL_Err_Ens_1 may be tested in the same manner to determine which regions of this prediction surface are predicted with confidence (low error), and which areas should be treated with caution. The results of this are shown in the 3D surface plot of Figure 7.3d, which plots error *versus* BD and %OM.

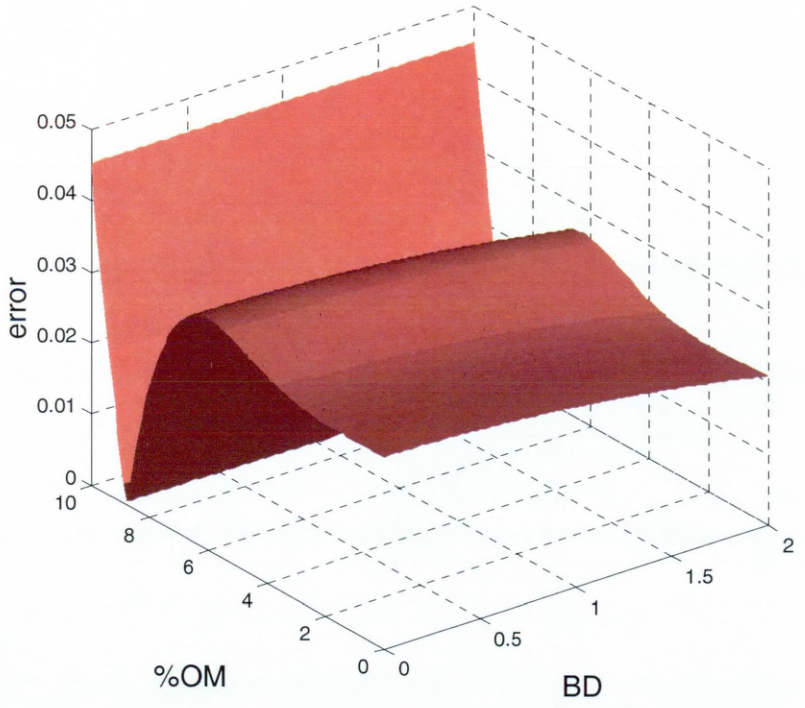



*Figure 7.3d: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble HCL_Ens_3 and error ensemble HCL_Err_Ens_1*

Figure 7.3d shows that there is a smooth relationship between prediction error, BD and %OM. Note that the errors are greatest when %OM is large (with the exception of the anomalous constriction at %OM = 9). This is to be expected, since the range of %OM for M soils ([0, 9.4], Tables 5.2c and 5.2e, and Figure 6.2j) with a

mean of 1.6 (and a median of 1.2) indicates that the distribution for %OM is strongly right-skewed (Section 5.2.4). Thus, it is expected that confidence should be highest when %OM is small, and this is indeed what can be seen in Figure 7.3d. In addition, since the effective range of BD for M soils (Tables 5.2c and 5.2e, and Figure 6.2i) is [0.76, 1.97], with a mean of 1.5, this indicates that the distribution of BD is left-skewed (Section 5.2.4). Therefore, confidence in the prediction of $\theta_{FC}$ should be highest when BD is large. Inspecting Figure 7.3d it can be seen that confidence is highest when BD is high and %OM is low, and this matches well to what should be expected.

In order to assess how well this error surface matches with the error surface that would be predicted by the Bootstrapping method, the same procedure has been carried out on TEST 5 and the results plotted as Figure 7.3e.

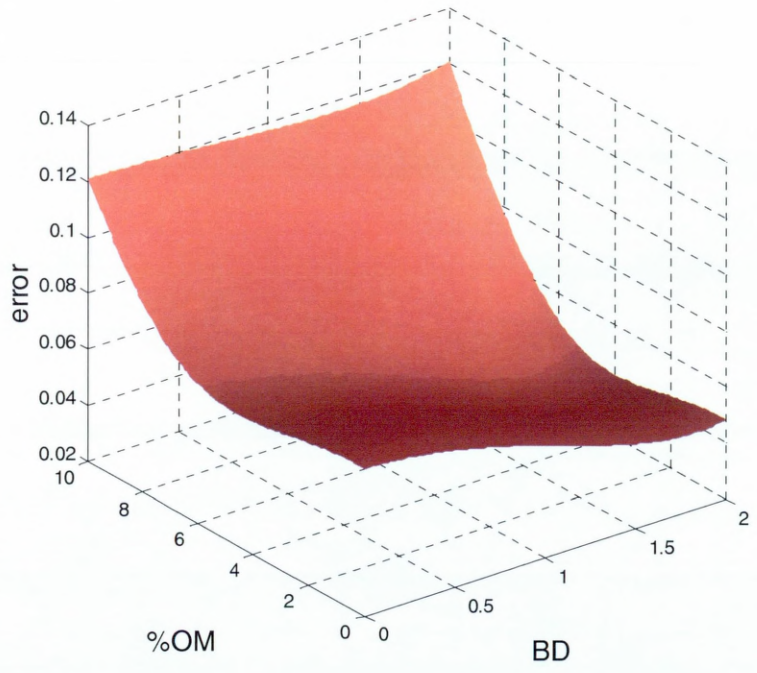



*Figure 7.3e: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble HCL_Ens_3 and the Bootstrapping method of error prediction*

The Bootstrapping method of error prediction shows that there is a smooth relationship between prediction error, BD and %OM. Note that the region of smallest error (and therefore highest confidence) is when BD is large and %OM is small,

corresponding to the effective ranges of [0.76, 1.97] and [0, 3] for BD and %OM, respectively.

## Chapter Conclusions

The same procedures that were used in Chapter 6 to construct a continuous PTF have been used to construct a class PTF. Using only the M soil texture class, I built a HYPRES class ANN-PTF ensemble of $\theta_{FC}$, using ten ANN ensemble members, each trained with 100 soil horizons Bootstrapped from the M texture class.

Test results show that the ensemble is stable and unbiased, and is insensitive to the loss of any individual input parameter when sensitivity analyses are performed. Omitting BD & %OM from the model results in a decrease in the ability of the class PTF to predict $\theta_{FC}$. Models that have no soil textural component (i.e. only BD and %OM are present) have significantly decreased predictive ability.

The two competing methods of producing error bars on individual predictions (Bootstrapping and error ensemble) have both been employed, and found to produce similarly confident results in the region of high BD and low %OM, as expected, albeit that the error ensemble method did produce some anomalous effects.

# Chapter 8

# Testing Models on Alternative Databases

Chapters 6 and 7 detailed the procedures to construct optimised class and continuous ANN-PTF ensembles of water retention at field capacity with supplementary models for error prediction, and those models were implemented on data from the HYPRES database. In Chapter 8 those models are used to construct class and continuous PTFs using data from the USDA-NRCS database. Firstly, the various selection criteria detailed in Chapter 5 will be applied to the available data, and the data analysed in precisely the manner discussed in Chapter 5. The class and continuous PTFs are then constructed, and sensitivity analyses are conducted on the ensemble to determine which of the input parameters are of most importance to such a model. In addition, the two methods of producing error bars on individual predictions detailed in Chapter 6 are employed here, and the results of these two competing methods are discussed in conjunction with the results of statistical analyses conducted in Section 8.1.

## 8.1 Results of Statistical Analyses of USDA-NRCS Database

In order to test the ANN-PTF ensemble of soil water retention at FC, the USDA-NRCS pedon database (USDA Natural Resource Conservation Service 1994) has been made available. This database has a different structural basis to that of HYPRES, however, the critical information that is required for the model (inputs of SSC, BD and %OM, and output of $\theta_{-33kPa}$) is stored within. For the sake of simplicity (Occam's razor), $\theta_{-33kPa}$ in the USDA-NRCS database is regarded as being equivalent to $\theta_{FC}$ in HYPRES, and will subsequently be termed as $\theta_{FC}$.

In HYPRES, each soil horizon is labelled as one of five FAO-defined mineral texture classes (FAO 1990). The USDA-NRCS database employs no such labelling distinction, and hence, each soil horizon was filtered through a purposely-written computer program to distinguish between texture classes. Each soil horizon was assessed in turn, determining whether, according to the definitions in Table 5.2d, it belonged to the C, M, MF, F or VF texture classes, and labelled accordingly.

On completion, each of the five FAO (1990) texture classes was examined with the same statistical tests as were used in Chapter 5, prior to using them in the modelling process.

Recall that model selection criteria demand that all horizon duplicates be deleted, and that only mineral soils (%OM < 10) shall be used. These criteria have been applied to USDA-NRCS, and are detailed by texture class in Table 8.1a. The distribution of these datapoints on the soil textural triangle are shown in Figure 8.1a.

| | Texture Class | | | | | |
|---|---|---|---|---|---|---|
| | **C** | **M** | **MF** | **F** | **VF** | **ALL** |
| Total number of soil horizons selected | 1967 | 5833 | 1651 | 1500 | 293 | 11244 |
| Number where %OM > 10 (deleted) | 17 | 95 | 9 | 11 | 2 | 134 |
| Number of duplicates (deleted) | 0 | 0 | 0 | 0 | 0 | 0 |
| **Total number remaining** | **1950** | **5738** | **1642** | **1489** | **291** | **11110** |
| Percentage of total remaining | 18 | 51 | 15 | 13 | 3 | 100 |

*Table 8.1a: Textural breakdown of the data selected for modelling*

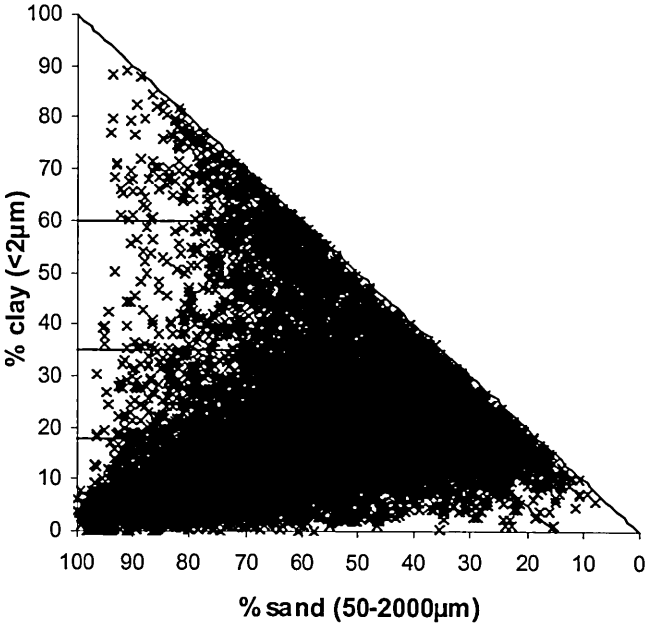



*Figure 8.1a: Class distribution of the data selected according to the selection criteria*

Of the 11244 soil horizons selected prior to application of the selection criteria, 134 of these were deleted (134 had %OM > 10, and there were no duplicates), leaving 11110 viable soil horizons for the modelling process.

Approximately one-half belong to the M texture class, whilst a little under one-half are distributed more or less equally across the C, MF and F texture classes. The remaining 3% belong to the little-represented VF texture class.

### 8.1.1 Do the Soil Textural Components Sum to 100?

It is expected that $\Sigma(SSC) = 100$ for all samples, and so each of the individual textural components of each soil horizons was summed and rounded to the nearest integer (Table 8.1b).

| | C | M | MF | F | VF | ALL |
|---|---|---|---|---|---|---|
| Total number of soil horizons | 1950 | 5738 | 1642 | 1489 | 291 | 11110 |
| Number of horizons for which $\Sigma(SSC) = 100$ | 1946 | 5725 | 1635 | 1487 | 289 | 11082 |
| Number of horizons for which $\Sigma(SSC) \neq 100$ | 4 | 13 | 7 | 2 | 2 | 28 |
| $\Sigma$ (SSC) when SSC $\neq 100$ | 104<br>101<br>99<br>101 | 101<br>96<br>101<br>96<br>80<br>70<br>99<br>99<br>101<br>68<br>105<br>101<br>101 | 78<br>98<br>99<br>45<br>98<br>103<br>101 | 105<br>98 | 120<br>101 | |

*Table 8.1b: Sum of proportions of SSC, and details where $\Sigma(SSC) \neq 100$*

Of the 11110 soil horizons that were selected for the modelling process, a total of 28 were found to have a mineral content that did not sum to 100. C, F and VF had four, two and two and M and MF had 13 and seven deviant horizons, respectively.

Of those samples that were deviant, 22 of the 28 were in the range [95, 105]. This may be considered reasonable since these are all within 5% tolerance, however, the six samples for which $\Sigma(SSC)$ = 70, 80, 68, 78, 45 and 120 certainly cannot be considered acceptable. Nonetheless, these deviant horizons have been included in the modelling process, as they were in the models implemented on HYPRES data.

### 8.1.2 Anderson-Darling Test for Normality

Recall that this normality test provides two goodness-of-fit measures to help assess how the distribution fits the data set. If $A^2 > 0.753$, there is 95% confidence that the distribution is non-normal. If $P > 0.05$, there is 95% confidence that the data set is normally distributed.

Table 8.1c show the results of the Anderson-Darling normality test when applied to various sub-categories of the dataset. Each soil texture class is composed of the five model input parameters (SSC, BD and %OM) and the one output parameter ($\theta_{FC}$).

All of these distributions are clearly highly non-normal, having values of $A^2$ that are greater than 0.753, and values of P that are very much smaller than 0.05.

| | %Cl | | %Si | | %Sa | |
|---|---|---|---|---|---|---|
| | $A^2$ | P | $A^2$ | P | $A^2$ | P |
| **C** | 12.9 | 0.000 | 6.4 | 0.000 | 23.4 | 0.000 |
| **M** | 16.1 | 0.000 | 9.3 | 0.000 | 46.2 | 0.000 |
| **MF** | 10.5 | 0.000 | 1.6 | 0.000 | 21.5 | 0.000 |
| **F** | 24.6 | 0.000 | 2.4 | 0.000 | 37.7 | 0.000 |
| **VF** | 5.8 | 0.000 | 2.4 | 0.000 | 11.8 | 0.000 |

| | BD | | %OM | | $\theta_{FC}$ | |
|---|---|---|---|---|---|---|
| | $A^2$ | P | $A^2$ | P | $A^2$ | P |
| **C** | 55.6 | 0.000 | 195.2 | 0.000 | 42.4 | 0.000 |
| **M** | 54.3 | 0.000 | 333.3 | 0.000 | 19.8 | 0.000 |
| **MF** | 25.9 | 0.000 | 64.9 | 0.000 | 13.8 | 0.000 |
| **F** | 18.0 | 0.000 | 60.7 | 0.000 | 5.2 | 0.000 |
| **VF** | 2.9 | 0.000 | 13.3 | 0.000 | 1.26 | 0.003 |

*Table 8.1c: Results of normality testing on the five input and one output parameters chosen for modelling*

### 8.1.3 Minimum and Maximum

Considering that the distributions of the individual parameters are highly non-normal, it may prove to be useful to determine their ranges, detailed by texture class, Table 8.1d.

| | %Cl | | %Si | | %Sa | |
|---|---|---|---|---|---|---|
| | **Min** | **Max** | **Min** | **Max** | **Min** | **Max** |
| **C** | 0 | 17.6 | 0 | 33.3 | 65.1 | 98.3 |
| **M** | 0 | 35 | 3.4 | 81.6 | 0.2 | 78 |
| **MF** | 1 | 34.9 | 22.1 | 92.2 | 0.5 | 14.9 |
| **F** | 35.1 | 59.9 | 4.5 | 63.7 | 0.3 | 58.8 |
| **VF** | 60.1 | 89.2 | 5.9 | 38.4 | 0.3 | 31 |

| | BD | | %OM | | $\theta_{FC}$ | |
|---|---|---|---|---|---|---|
| | **Min** | **Max** | **Min** | **Max** | **Min** | **Max** |
| **C** | 0.2 | 2.16 | 0.01 | 9.51 | 0.019 | 0.737 |
| **M** | 0.16 | 2.01 | 0.02 | 9.97 | 0.042 | 0.733 |
| **MF** | 0.13 | 2.13 | 0.03 | 9.6 | 0.082 | 0.618 |
| **F** | 0.13 | 1.77 | 0.1 | 9.44 | 0.047 | 0.724 |
| **VF** | 0.25 | 1.54 | 0.18 | 8.94 | 0.293 | 0.791 |

*Table 8.1d: Minimum and maximum values of the categorised parameters*

Comparing Tables 5.2d and 8.1d, the ranges of all texture classes fall within their FAO (1990) defined ranges.

### 8.1.4 Mean and Median

Now that we know the ranges of each of the categories of parameters, determining the mean and median of such will allow the investigator to understand where the average values lie, Table 8.1e.

| | %Cl | | %Si | | %Sa | |
|---|---|---|---|---|---|---|
| | **Mean** | **Median** | **Mean** | **Median** | **Mean** | **Median** |
| **C** | 6.86 | 6.50 | 15.92 | 16.30 | 77.22 | 75.90 |
| **M** | 17.39 | 16.80 | 42.27 | 41.50 | 40.32 | 40.60 |
| **MF** | 23.24 | 23.30 | 68.95 | 68.60 | 7.76 | 7.30 |
| **F** | 45.60 | 44.70 | 40.11 | 40.40 | 14.29 | 11.70 |
| **VF** | 67.65 | 66.50 | 24.94 | 26.50 | 7.48 | 5.80 |

| | BD | | %OM | | $\theta_{FC}$ | |
|---|---|---|---|---|---|---|
| | **Mean** | **Median** | **Mean** | **Median** | **Mean** | **Median** |
| **C** | 1.439 | 1.500 | 1.07 | 0.630 | 0.182 | 0.167 |
| **M** | 1.313 | 1.350 | 1.88 | 1.350 | 0.294 | 0.290 |
| **MF** | 1.339 | 1.360 | 1.74 | 1.490 | 0.346 | 0.346 |
| **F** | 1.256 | 1.280 | 1.96 | 1.580 | 0.392 | 0.390 |
| **VF** | 1.132 | 1.150 | 1.89 | 1.440 | 0.458 | 0.457 |

*Table 8.1e: Mean and median of the categorised parameters*

On average, as the %Cl increases, the average $\theta_{FC}$ increases, and also that as the average BD increases, the average $\theta_{FC}$ decreases.

Comparing each of the categorised parameter values for the mean and median (Table 8.1e) with their respective ranges (Table 8.1d) allows an understanding of whether their distributions are right- or left-skewed. To do this, I calculated the distance from the minimum to the maximum value and determined, as an integer percentage, where along this line the mean and median lie (Table 8.1f). The mean results have also been plotted, Figure 8.1a (the median has been omitted, since there is no significant difference between the mean and median values for each of the categorised parameters).

| | %Cl | | %Si | | %Sa | |
|---|---|---|---|---|---|---|
| | **Mean** | **Median** | **Mean** | **Median** | **Mean** | **Median** |
| **C** | 39 | 37 | 48 | 49 | 37 | 33 |
| **M** | 50 | 48 | 50 | 49 | 52 | 52 |
| **MF** | 66 | 66 | 67 | 66 | 50 | 47 |
| **F** | 42 | 39 | 60 | 61 | 24 | 19 |
| **VF** | 26 | 22 | 59 | 63 | 23 | 18 |

| | BD | | %OM | | $\theta_{FC}$ | |
|---|---|---|---|---|---|---|
| | **Mean** | **Median** | **Mean** | **Median** | **Mean** | **Median** |
| **C** | 63 | 66 | 11 | 7 | 23 | 21 |
| **M** | 62 | 64 | 19 | 13 | 36 | 36 |
| **MF** | 60 | 62 | 18 | 15 | 49 | 49 |
| **F** | 69 | 70 | 20 | 16 | 51 | 51 |
| **VF** | 68 | 70 | 20 | 14 | 33 | 33 |

*Table 8.1f: Relative position of the mean and median values of categorised parameters, as a percentage of the distance between minimum and maximum*

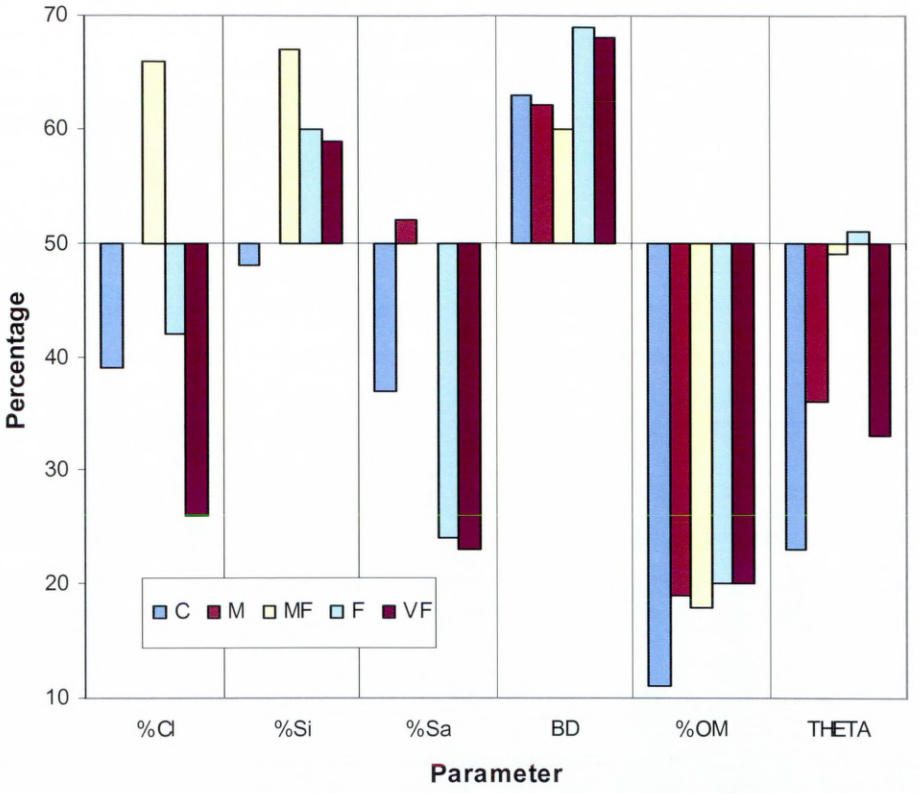



*Figure 8.1a: Averages of SSC, plotted as a percentage of the respective distances from their minimum to maximum values*

An analysis of Table 8.1f and Figure 8.1a show that:

- BD is strongly left-skewed
- $\theta_{FC}$ is mostly right-skewed
- %OM is strongly right-skewed
- %Cl is right-skewed for C, F and VF soils, but left-skewed for MF soils
- %Si is left-skewed for high proportions of clay
- %Sa is mostly right-skewed, also, as %Cl increases, distribution becomes increasingly right-skewed

### 8.1.5 Standard Deviation and Coefficient of Variation

The σ and CV of the categorised parameters have been calculated, and are shown in Table 8.1g.

| | %Cl | | %Si | | %Sa | |
|---|---|---|---|---|---|---|
| | σ | CV | σ | CV | σ | CV |
| **C** | 3.79 | 55.16 | 6.95 | 43.63 | 8.13 | 10.53 |
| **M** | 7.68 | 44.15 | 12.82 | 30.33 | 14.12 | 35.03 |
| **MF** | 6.82 | 29.34 | 7.65 | 11.09 | 3.87 | 49.88 |
| **F** | 7.23 | 15.86 | 10.92 | 27.23 | 10.56 | 73.90 |
| **VF** | 5.99 | 8.85 | 7.84 | 31.41 | 6.52 | 87.19 |

| | BD | | %OM | | $\theta_{FC}$ | |
|---|---|---|---|---|---|---|
| | σ | CV | σ | CV | σ | CV |
| **C** | 0.26 | 18.11 | 1.30 | 122.07 | 0.09 | 50.09 |
| **M** | 0.25 | 19.02 | 1.65 | 87.80 | 0.08 | 26.05 |
| **MF** | 0.16 | 12.31 | 1.19 | 68.16 | 0.05 | 14.31 |
| **F** | 0.18 | 14.59 | 1.49 | 75.72 | 0.06 | 16.06 |
| **VF** | 0.17 | 15.00 | 1.43 | 75.77 | 0.07 | 15.65 |

*Table 8.1g: σ and CV of the categorised parameters*

Recall that, in general, the larger σ or CV, the higher is the variability of the values about the mean. In particular, a small value for the σ and a high value for the CV indicate that there is a tight spread about the mean, but that those values are highly variable. Additionally, the main features of Table 8.1g are that:

- CV increases as %Cl decreases
- CV increases as %Sa decreases, in particular, the variability of %Sa in F and VF texture classes is very high
- Variability of %OM is very high
- As %Cl is decreased, the CV of %Cl and $\theta_{FC}$ both increase
- As %Sa is increased, the CV of %Sa decreases, and the CV of $\theta_{FC}$ increases

## 8.2 USDA-NRCS Continuous PTF

### 8.2.1 Results and Discussion

Using the continuous ANN-PTF ensemble method previously constructed in Chapter 6, the model has been re-trained using data from the USDA-NRCS database.

All training parameters remained unchanged, including the number of soil horizons per member (25 from each texture class) and number of member ANNs in the ensemble (ten). Ten ensembles were trained in total, and the RMSE results of testing (all tests) with the USDA-NRCS dataset are shown and detailed by texture class and overall (Table 8.2a), denoted UCO_Ens_1, UCO_Ens_2, ..., UCO_Ens_10, where UCO denotes that the model is a USDA-NRCS Continuous PTF.

| | RMSE | | | | | |
|---|---|---|---|---|---|---|
| | **C** | **M** | **MF** | **F** | **VF** | **ALL** |
| **UCO_Ens_1** | 0.072 | 0.063 | 0.050 | 0.058 | 0.066 | 0.063 |
| **UCO_Ens_2** | 0.074 | 0.066 | 0.046 | 0.058 | 0.065 | 0.064 |
| **UCO_Ens_3** | 0.072 | 0.065 | 0.047 | 0.055 | 0.066 | 0.063 |
| **UCO_Ens_4** | 0.072 | 0.066 | 0.049 | 0.057 | 0.066 | 0.064 |
| **UCO_Ens_5** | 0.076 | 0.064 | 0.046 | 0.056 | 0.068 | 0.063 |
| **UCO_Ens_6** | 0.077 | 0.065 | 0.051 | 0.055 | 0.068 | 0.065 |
| **UCO_Ens_7** | 0.073 | 0.066 | 0.046 | 0.056 | 0.067 | 0.063 |
| **UCO_Ens_8** | 0.075 | 0.066 | 0.047 | 0.055 | 0.063 | 0.064 |
| **UCO_Ens_9** | 0.075 | 0.065 | 0.047 | 0.056 | 0.067 | 0.063 |
| **UCO_Ens_10** | 0.074 | 0.066 | 0.047 | 0.056 | 0.067 | 0.064 |

*Table 8.2a: RMSEs of 10 ensembles, detailed by texture class and overall*

The results show a remarkable stability. The ensembles vary little in terms of their predictive abilities, overall and for each texture class.

Table 8.2b shows the ME results (all tests) of each of the ten ensembles, shown to four decimal places, and detailed by texture class and overall.

| | ME | | | | | |
|---|---|---|---|---|---|---|
| | **C** | **M** | **MF** | **F** | **VF** | **ALL** |
| **UCO_Ens_1** | -0.0017 | 0.0029 | -0.0042 | -0.0054 | 0.0066 | $1.98 \times 10^{-5}$ |
| **UCO_Ens_2** | -0.0040 | -0.0052 | -0.0049 | -0.0032 | -0.0026 | -0.0046 |
| **UCO_Ens_3** | 0.0020 | -0.0019 | 0.0050 | -0.0036 | 0.0023 | -0.0003 |
| **UCO_Ens_4** | 0.0003 | -0.0036 | 0.0029 | -0.0039 | 0.0015 | -0.0019 |
| **UCO_Ens_5** | 0.0031 | 0.0023 | 0.0016 | 0.0030 | -0.0008 | 0.0024 |
| **UCO_Ens_6** | -0.0137 | -0.0018 | 0.0008 | 0.0073 | 0.0144 | -0.0018 |
| **UCO_Ens_7** | -0.0036 | 0.0051 | 0.0026 | 0.0057 | 0.0026 | 0.0032 |
| **UCO_Ens_8** | -0.0101 | -0.0066 | -0.0011 | 0.0062 | 0.0076 | -0.0043 |
| **UCO_Ens_9** | -0.0090 | -0.0033 | -0.0020 | 0.0034 | 0.0127 | -0.0028 |
| **UCO_Ens_10** | -0.0004 | -0.0010 | 0.0015 | 0.0005 | 0.0078 | -0.0001 |

*Table 8.2b: MEs of 10 ensembles, detailed by texture class and overall*

Here, all MEs are at least one order of magnitude smaller than the corresponding RMSE results, indicating that the bias in each of the models is very small. It is also encouraging to note that the sign of the ME is not consistent, in other words, $\theta_{FC}$ is neither consistently under- nor over-estimated. This is in contrast to results published by Nemes *et al.* (2003) who found that water retention was always under-estimated in their models, indicating a significant bias towards datasets having small %OM.

Table 8.2c shows the average results, both RMSE and ME, of the ten ensembles, detailed by texture class and overall.

| | C | M | MF | F | VF | ALL |
|---|---|---|---|---|---|---|
| **Number of tests** | 1950 | 5738 | 1642 | 1489 | 291 | 11110 |
| **RMSE** | 0.074 | 0.065 | 0.048 | 0.056 | 0.066 | 0.064 |
| **ME** | -0.0037 | -0.0013 | 0.0002 | 0.0010 | 0.0052 | -0.0010 |

*Table 8.2c: RMSE and ME results, averaged over 10 ensemble models*

The RMSEs for M and VF texture classes are very similar, whilst the RMSEs of F and MF are significantly better, and the RMSE of C soil horizons are appreciably worse. Note again that all MEs are at least one order of magnitude smaller than the corresponding RMSE results.

As with the modelling of the HYPRES data, sensitivity analyses have been conducted on the USDA-NRCS model. Each of the five input parameters were omitted in turn from the ensemble model, also SSC was omitted, and BD & %OM, and then each of these re-trained and re-tested in precisely the same way as previously. This procedure is repeated 10 times, and the results averaged. The RMSE results of testing are shown in Table 8.2d, and detailed by texture class. The results are also shown graphically in Figure 8.2a. Additionally, the RMSE results of the ensemble with no parameter omissions are repeated for comparative purposes (results averaged over 10 procedural repeats). The ME results varied very little from model to model and are not be shown here.

Also, the individual ensemble results of the SSC-omitted models will be detailed in Tables 8.2e (RMSE) and 8.2f (ME).

| Input parameters used | | | | | RMSE | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| %Cl | %Si | %Sa | BD | %OM | C | M | MF | F | VF | ALL |
| ✓ | ✓ | ✓ | ✓ | ✓ | 0.074 | 0.065 | 0.048 | 0.056 | 0.066 | 0.064 |
| | ✓ | ✓ | ✓ | ✓ | 0.073 | 0.065 | 0.047 | 0.056 | 0.064 | 0.063 |
| ✓ | | ✓ | ✓ | ✓ | 0.075 | 0.065 | 0.047 | 0.056 | 0.065 | 0.064 |
| ✓ | ✓ | | ✓ | ✓ | 0.073 | 0.065 | 0.048 | 0.056 | 0.067 | 0.063 |
| ✓ | ✓ | ✓ | | ✓ | 0.076 | 0.065 | 0.046 | 0.057 | 0.068 | 0.064 |
| ✓ | ✓ | ✓ | ✓ | | 0.074 | 0.067 | 0.049 | 0.056 | 0.066 | 0.064 |
| ✓ | ✓ | ✓ | | | 0.083 | 0.070 | 0.048 | 0.056 | 0.069 | 0.068 |
| | | | ✓ | ✓ | See Tables 8.2e and 8.2f | | | | | |

*Table 8.2d: RMSE results of sensitivity analyses of the ensemble method, averaged over 10 models*

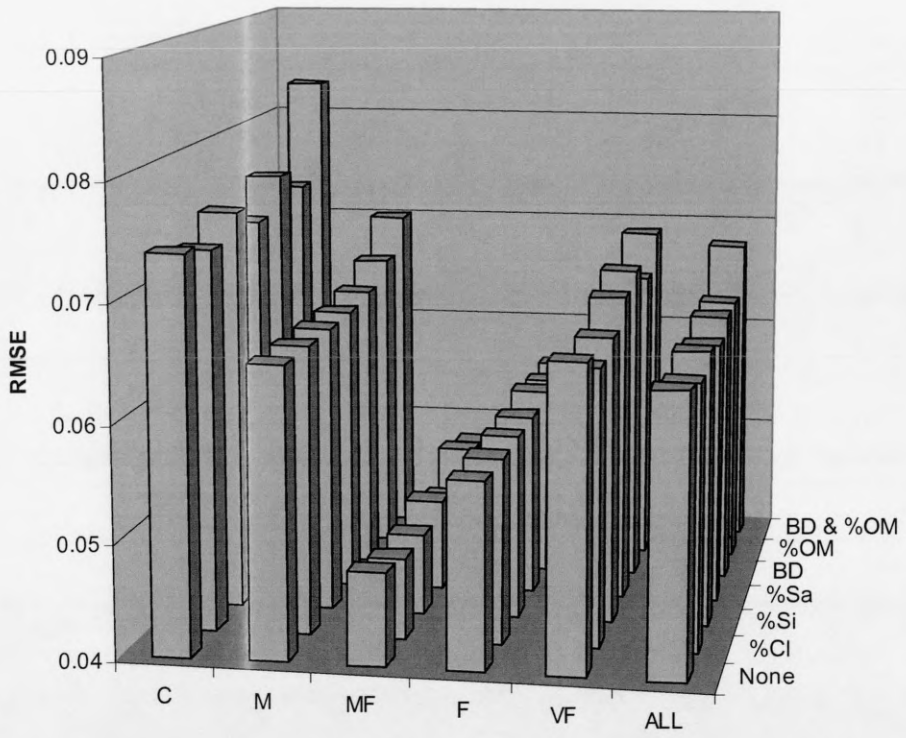



*Figure 8.2a: Graphical representation of sensitivity analyses of the ensemble method, averaged over 10 models*

Interestingly, omitting any of the input parameters as a single omission does not affect the model in any way, either by texture class or overall. Omitting BD & %OM

results in a small decrease in predictive ability for the M and VF texture classes, and results in a significant decrease in predictive ability for C soil horizons.

| | RMSE | | | | | |
|---|---|---|---|---|---|---|
| | C | M | MF | F | VF | ALL |
| UCO_Ens_1 | 0.876 | 1.456 | 0.752 | 0.943 | 0.879 | 1.205 |
| UCO_Ens_2 | 0.180 | 0.255 | 0.136 | 0.161 | 0.116 | 0.214 |
| UCO_Ens_3 | 0.913 | 0.611 | 0.302 | 0.202 | 0.140 | 0.599 |
| UCO_Ens_4 | 1.373 | 1.754 | 0.952 | 1.058 | 1.064 | 1.495 |
| UCO_Ens_5 | 0.173 | 0.229 | 0.126 | 0.152 | 0.115 | 0.195 |
| UCO_Ens_6 | 0.133 | 0.094 | 0.064 | 0.080 | 0.106 | 0.097 |
| UCO_Ens_7 | 0.138 | 0.134 | 0.082 | 0.096 | 0.092 | 0.123 |
| UCO_Ens_8 | 0.145 | 0.162 | 0.090 | 0.109 | 0.105 | 0.142 |
| UCO_Ens_9 | 0.263 | 0.400 | 0.199 | 0.280 | 0.227 | 0.335 |
| UCO_Ens_10 | 0.519 | 0.725 | 0.418 | 0.680 | 0.565 | 0.644 |

*Table 8.2e: RMSEs of 10 ensembles (SSC omitted), detailed by texture class and overall*

Considering the results shown in Table 8.2e, it can be seen that the ensemble results show a great variation in predictive abilities. The best ensemble (SSC omitted) is poor when compared to the ensemble with no omissions, whilst the worst ensemble (SSC omitted) appears to fail completely to match the inputs to the outputs.

| | ME | | | | | |
|---|---|---|---|---|---|---|
| | C | M | MF | F | VF | ALL |
| UCO_Ens_1 | -0.1454 | -0.1741 | -0.0226 | -0.0230 | 0.0219 | -0.1213 |
| UCO_Ens_2 | -0.1030 | -0.0546 | 0.0131 | 0.0316 | 0.0625 | -0.0385 |
| UCO_Ens_3 | -0.0144 | -0.0345 | 0.0147 | 0.0251 | 0.0576 | -0.0133 |
| UCO_Ens_4 | 0.0399 | 0.1577 | 0.0734 | 0.1167 | 0.1324 | 0.1184 |
| UCO_Ens_5 | -0.1048 | -0.0559 | 0.0103 | 0.0293 | 0.0601 | -0.0402 |
| UCO_Ens_6 | -0.1011 | -0.0314 | 0.0152 | 0.0451 | 0.0780 | -0.0236 |
| UCO_Ens_7 | -0.1000 | -0.0413 | 0.0156 | 0.0360 | 0.0650 | -0.0301 |
| UCO_Ens_8 | -0.1036 | -0.0448 | 0.0135 | 0.0374 | 0.0783 | -0.0323 |
| UCO_Ens_9 | -0.1115 | -0.0781 | 0.0039 | 0.0143 | 0.0458 | -0.0562 |
| UCO_Ens_10 | -0.1568 | -0.1616 | -0.0305 | -0.0615 | -0.0095 | -0.1240 |

*Table 8.2f: MEs of 10 ensembles (SSC omitted), detailed by texture class and overall*

Considering Table 8.2f, the MEs in many cases are large enough to be regarded as significant, and the sign of the bias shows that the $\theta_{FC}$ of C and M soil horizons are mostly over-estimated, whilst MF, F and VF are mostly under-estimated (Figure 8.2b).

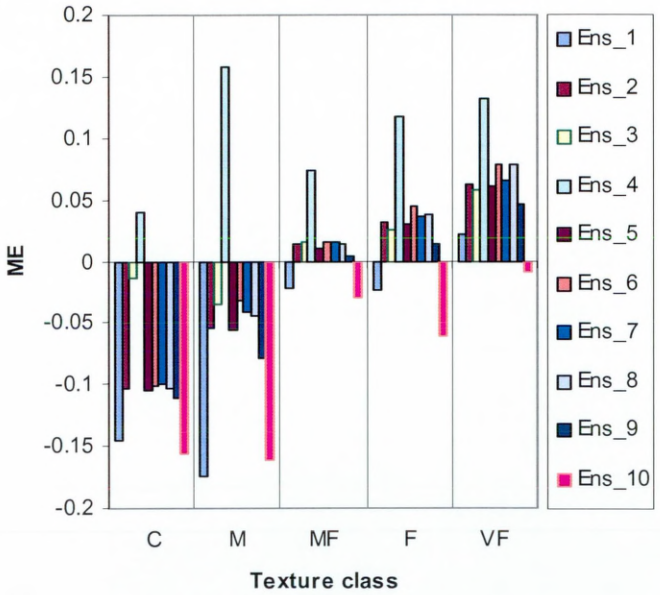



*Figure 8.2b: MEs of ten ensembles (SSC omitted), detailed by texture class*

Figure 8.2b shows that, for most of the SSC-omitted ensembles, as the %Cl increases (reading from C to VF) the ensembles progressively under-estimate $\theta_{FC}$.

Thus, although omitting parameters individually from the model has no effect on the ensemble for the USDA-NRCS pedon database, depriving it of all textural information results in an unreliable model.

Figure 8.2c shows how the ensemble-predicted $\theta_{FC}$ values correlate with their expected values, for UCO_Ens_1, detailed by texture class and overall.

There is good agreement between expected and ensemble-predicted values of $\theta_{FC}$ for all texture classes and overall, with relatively close fitting to the expected 1:1 correspondence lines. There does not appear to be consistent under or overestimations for any of the texture classes or overall.

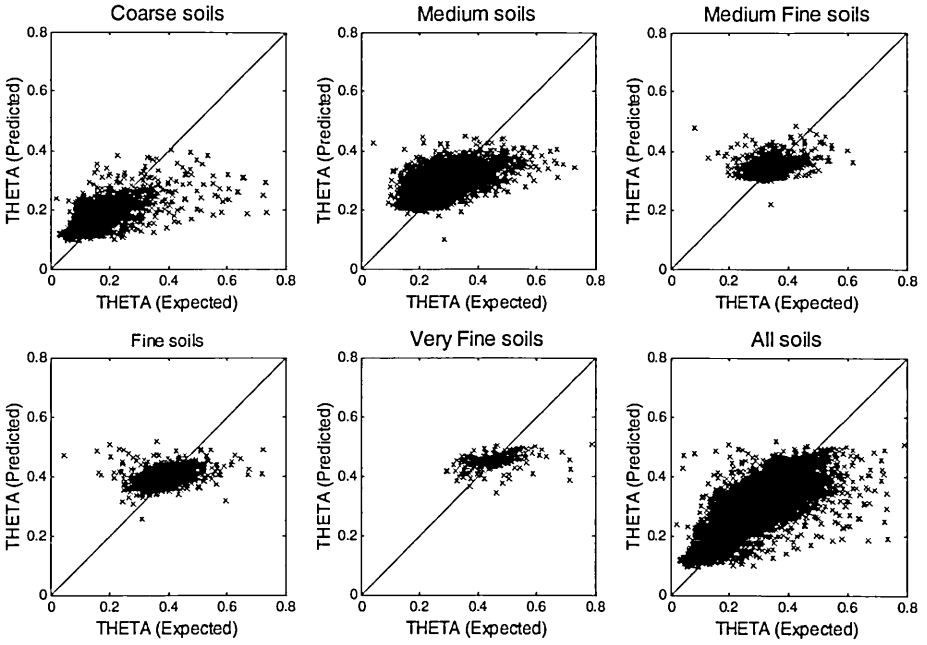



*Figure 8.2c: Comparison of expected and ensemble-predicted $\theta_{FC}$ for ensemble model UCO_Ens_1, detailed by texture class and overall*

To visualise the effect of varying BD on predictions of $\theta_{FC}$ for the USDA-NRCS continuous ANN-PTF ensemble, the same five horizons as those of Table 6.2h were used to re-test UCO_Ens_1. BD was varied in increments of 0.01 in the range [0, 2], UCO_Ens_1 was re-tested, and the results plotted as $\theta_{FC}$ *versus* BD. The results of such tests are shown graphically in Figure 8.2d, additionally, the BD has been plotted against measured values of $\theta_{FC}$ for all the USDA-NRCS M soil horizons used in the modelling process, and a linear regression line plotted.

As BD increases, $\theta_{FC}$ decreases non-linearly for all M soils tested. From Tables 8.1d and 8.1e, the range of BD for M soils is [0.16, 2.01], with a mean of 1.3 (median of 1.35), and the range of $\theta_{FC}$ for M soils is [0.04, 0.73], with a mean of 0.29. The regression line on Figure 8.2d shows that one would expect that as the BD is increased $\theta_{FC}$ should decrease, and this is exactly what is seen. It should be noted that there are no soil samples that have zero BD, since this would represent a soil that has zero density, i.e. a solid pore composed entirely of air, and thus, the model has no pedologic meaning for soils having BD < 0.16 (since the model would be extrapolating). Additionally, the gradient of the regression line in Figure 8.2d is not

significantly different from the gradients of the five test samples shown in the range [0.8, 2].

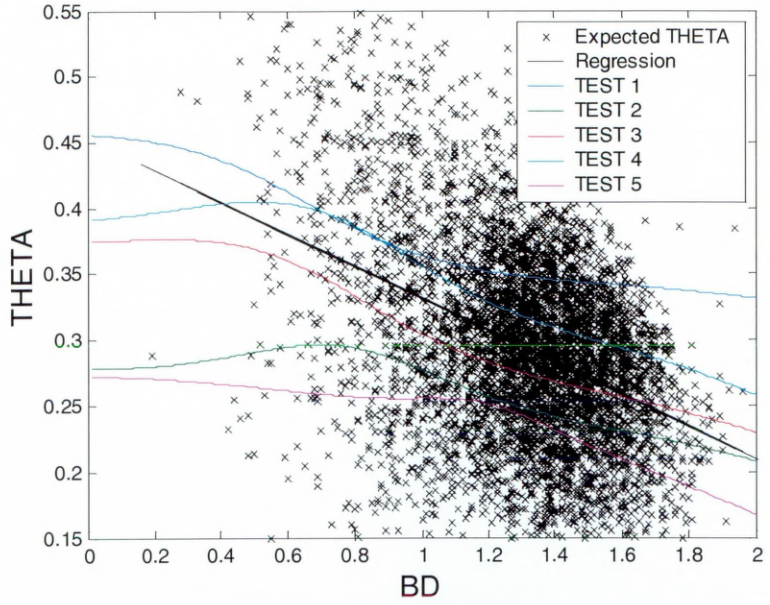



*Figure 8.2d: Effect of varying BD on five randomly chosen M soil horizons, with measured values of $\theta_{FC}$ and regression line*

This procedure has also been repeated with respect to %OM, and incremented by 0.1 in the range [0, 10]. The results of $\theta_{FC}$ *versus* %OM are shown in Figure 8.2e, additionally, the %OM has been plotted against measured values of $\theta_{FC}$ for all the M soils used in the modelling process, and a linear regression line plotted.

As %OM is increased, $\theta_{FC}$ increases for all M soils tested. From Tables 8.1d and 8.1e, the range of %OM for M soils is [0.02, 9.97], with a mean of 1.88 (and a median of 1.35), and the range of $\theta_{FC}$ for M soils is [0.04, 0.73], with a mean of 0.29. The regression line on Figure 8.2e shows that one would expect that as the %OM is increased $\theta_{FC}$ should increase, and this is exactly what is seen for all sample tests. It should be noted that the majority of %OM values for M soil horizons at $\theta_{FC}$ fall in the range [0, 3], and since there is much less data in the range [3, 10], it should be expected that the model produces more reliable predictions for %OM < 3. The gradient of the regression line in Figure 8.2e is not significantly different from the gradients of the five test samples shown.

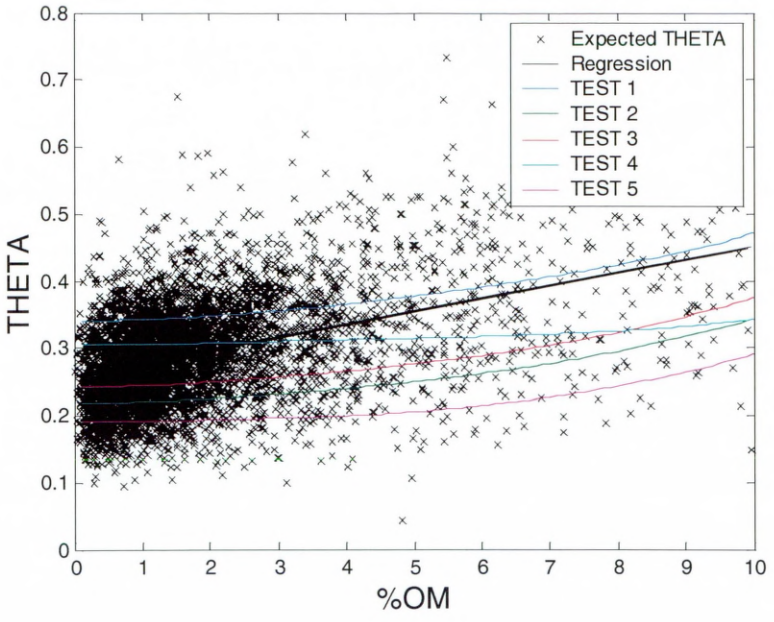



*Figure 8.2e: Effect of varying %OM on five randomly chosen M soil horizons, with measured values of $\theta_{FC}$ and regression line*

## 8.3 USDA-NRCS Class PTF

### 8.3.1 Results and Discussion

Chapter 7 detailed the protocol for constructing a class ANN-PTF ensemble of soil water retention at field capacity from the HYPRES dataset, and now these same procedures are used to construct a class ANN-PTF ensemble of water retention at field capacity for the USDA-NRCS dataset. Recall that the only difference in the procedures between class and continuous models is that in the continuous PTF, 25 soil horizons per ANN member were used from each texture class, whilst in the class PTF, 100 soil horizons from the M soil texture class were used.

Table 8.3a shows the RMSE and ME results (all tests) of each of the ten ensembles, denoted UCL_Ens_1, UCL_Ens_2, …, UCL_Ens_10, where UCL denotes that the model is a USDA-NRCS Class PTF.

| | M | |
|---|---|---|
| | **RMSE** | **ME** |
| **UCL_Ens_1** | 0.065 | 0.0038 |
| **UCL_Ens_2** | 0.066 | -0.0014 |
| **UCL_Ens_3** | 0.065 | 0.0022 |
| **UCL_Ens_4** | 0.065 | -0.0014 |
| **UCL_Ens_5** | 0.064 | 0.0022 |
| **UCL_Ens_6** | 0.065 | 0.0043 |
| **UCL_Ens_7** | 0.065 | -0.0036 |
| **UCL_Ens_8** | 0.067 | -0.0033 |
| **UCL_Ens_9** | 0.068 | -0.0004 |
| **UCL_Ens_10** | 0.067 | -0.0013 |

*Table 8.3a: RMSEs and MEs of 10 ensembles*

In Section 6.1.3 it was established that, for the amount of data that is available for these investigations, the RMSE results are only reliable to two decimal places, however, quoting to three decimal places demonstrates the stability of the ensembles, each varying little (RMSE) in terms of their predictive abilities.

Here, all MEs are at least one order of magnitude smaller than the corresponding RMSE results, indicating that the bias in each of the models is very small. Also note that the sign of the ME is not consistent, indicating that $\theta_{FC}$ is neither consistently under- nor over-estimated.

Table 8.3b shows the average results, both RMSE and ME, of the ten ensembles.

Figure 8.3a shows how the ensemble-predicted values correlate with the expected values of $\theta_{FC}$, for USDA-NRCS class ANN-PTF ensemble model UCL_Ens_5 of the M texture class.

| | M |
|---|---|
| **Number of tests** | 5738 |
| **RMSE** | 0.066 |
| **ME** | 0.0001 |

*Table 8.3b: RMSE and ME results, averaged over 10 ensemble models*

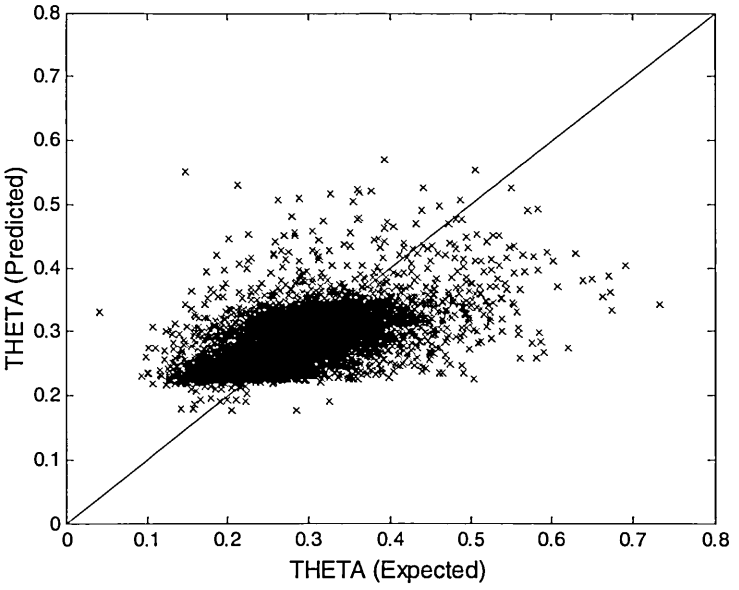



*Figure 8.3a: Comparison of expected and ensemble-predicted $\theta_{FC}$ for ensemble model UCL_Ens_5 of the M texture class*

There is a good correspondence between ensemble-predicted and expected values of $\theta_{FC}$ for M textured soil horizons, with most values falling on or near to the expected 1:1 correspondence. The model does not consistently under or overestimate $\theta_{FC}$, indicating that the model is unbiased.

A comparison of the continuous and class ensemble-predicted and expected values of $\theta_{FC}$ for the USDA-NRCS M texture class will be given in Section 9.2.

To visualise the effect of varying BD on predictions of $\theta_{FC}$ for the USDA-NRCS class ANN-PTF ensemble, the same five horizons as those of Table 6.2h were used to re-test UCL_Ens_5. BD was varied in increments of 0.01 in the range [0, 2], UCL_Ens_5 was re-tested, and the results plotted as $\theta_{FC}$ *versus* BD. The results of such tests are shown graphically in Figure 8.3b.

As BD increases, $\theta_{FC}$ decreases for all M soils tested, and the magnitude of the decrease is small and consistent. From Tables 8.1d and 8.1e, the range of BD for M soils is [0.16, 2.01], with a mean of 1.3 (median of 1.35), and the range of $\theta_{FC}$ for M soils is [0.04, 0.73], with a mean of 0.29. The regression line on Figure 8.3b shows that one would expect that as the BD is increased $\theta_{FC}$ should decrease, and this is

exactly what is seen. It should be noted that the model has no pedologic meaning for soils having BD < 0.16 (since the model would be extrapolating). Additionally, the gradient of the regression line in Figure 8.3b is quite different from the gradients of the five test samples shown.

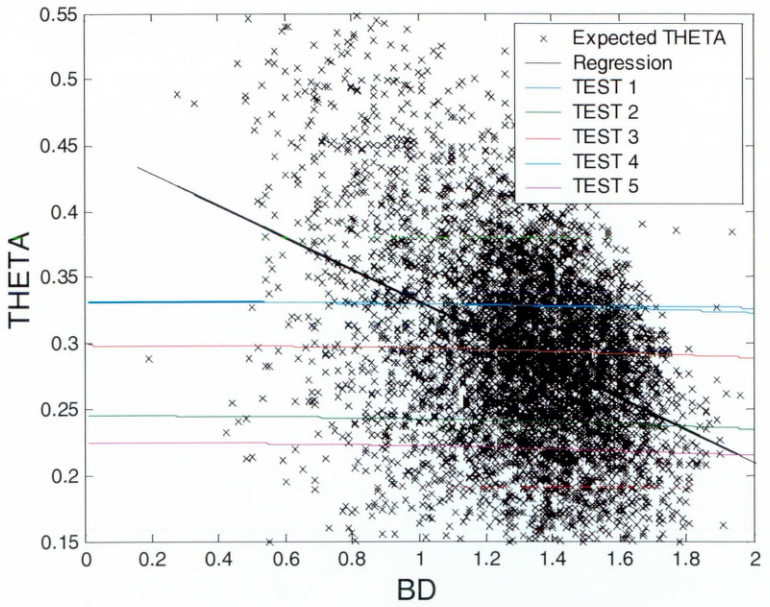



*Figure 8.3b: Effect of varying BD on five randomly chosen M soil horizons, with measured values of $\theta_{FC}$ and regression line*

This procedure has also been repeated with respect to %OM, and incremented by 0.1 in the range [0, 10]. The results of $\theta_{FC}$ *versus* %OM are shown in Figure 8.3c.

As %OM is increased, $\theta_{FC}$ increases for all M soils tested. From Tables 8.1d and 8.1e, the range of %OM for M soils is [0.02, 9.97], with a mean of 1.88 (and a median of 1.35), and the range of $\theta_{FC}$ for M soils is [0.04, 0.73], with a mean of 0.29. The regression line on Figure 8.3c shows that one would expect that as the %OM is increased $\theta_{FC}$ should increase, and this is exactly what is seen for all sample tests. It should be noted that the majority of %OM values for M soil horizons at $\theta_{FC}$ fall in the range [0, 3], and since there is much less data in the range [3, 10], it should be expected that the model produces more reliable predictions for %OM < 3. The gradient of the regression line in Figure 7.2c is not significantly different from the gradients of the five test samples shown.

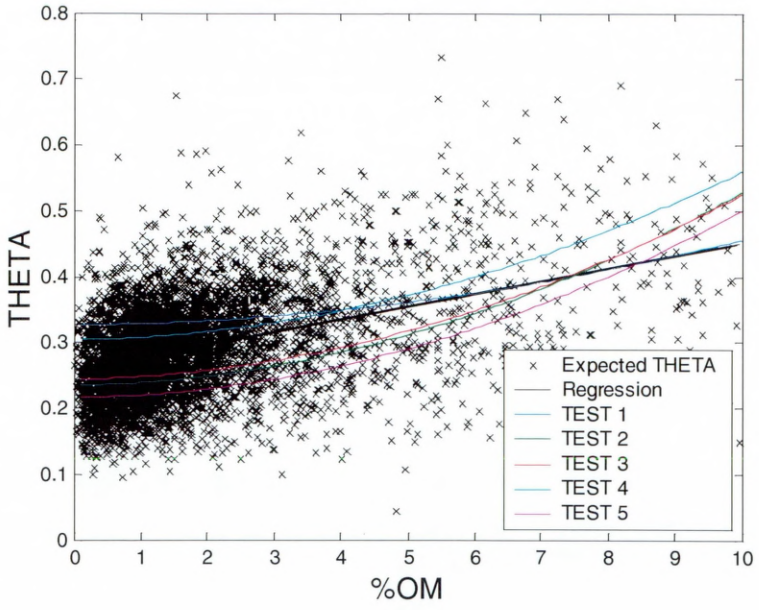



*Figure 8.3c: Effect of varying %OM on five randomly chosen M soil horizons, with measured values of $\theta_{FC}$ and regression line*

Sensitivity analyses have been conducted on the class PTF in the same manner as in Sections 6.2.5 and 7.2.2. Each of the five input parameters has been omitted in turn from the class PTF ensemble model, also SSC was omitted, and BD & %OM, and then each of these re-trained and re-tested in precisely the same way as in the ensemble method. This procedure is repeated 10 times, and the results averaged. The RMSE results of testing are shown in Table 8.3c. The results are also shown graphically in Figure 8.3d. Additionally, the RMSE results of the ensemble with no parameter omissions are repeated for comparative purposes (results averaged over 10 procedural repeats). The ME results varied very little from model to model and are not shown here.

| Input parameters used | | | | | RMSE |
|---|---|---|---|---|---|
| % Cl | % Si | % Sa | BD | %OM | M |
| ✓ | ✓ | ✓ | ✓ | ✓ | 0.066 |
| | ✓ | ✓ | ✓ | ✓ | 0.065 |
| ✓ | | ✓ | ✓ | ✓ | 0.066 |
| ✓ | ✓ | | ✓ | ✓ | 0.066 |
| ✓ | ✓ | ✓ | | ✓ | 0.065 |
| ✓ | ✓ | ✓ | ✓ | | 0.067 |
| ✓ | ✓ | ✓ | | | 0.069 |
| | | | ✓ | ✓ | 0.180 |

*Table 8.3c: RMSE results of sensitivity analyses of the ensemble method, averaged over 10 models*

Omitting an individual parameter has no significant effect on the class PTF as a whole. In terms of multiple parameter omissions, omitting BD & %OM has no significant effect, however, omitting SSC results in a significant deterioration in predictive ability of the ensemble.

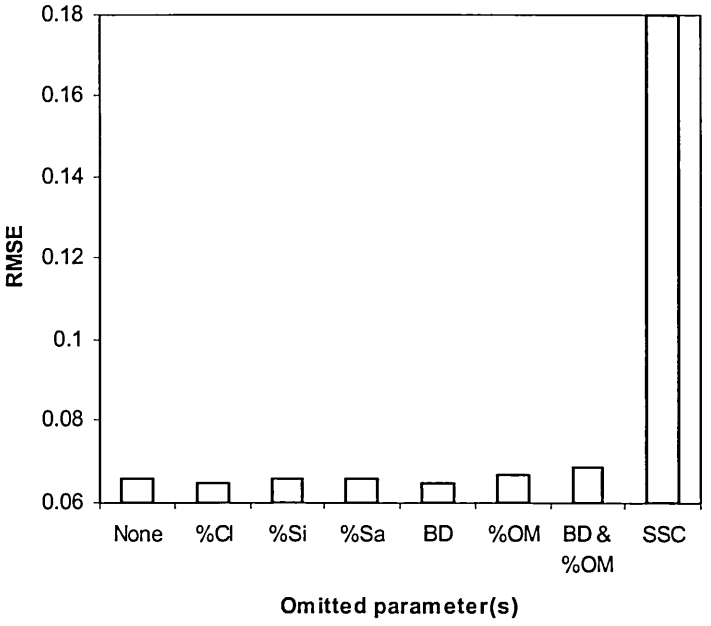



*Figure 8.3d: RMSEs of ensemble sensitivity analyses, averaged over 10 models*

## 8.4 Error Bar Modelling – Results and Discussion

### 8.4.1 Continuous PTF Error Bar Results

Analyses of the USDA-NRCS continuous PTF, error ensemble and Bootstrapping methods will be conducted in this section using M soil horizon TEST 5 (Table 6.2h), and results shown graphically as error bars and 3D error surfaces. These will enable a practical comparison of the two methods of error prediction, and will illustrate their comparative strengths and weaknesses. Additionally, the results of Section 8.2.1 will be discussed with reference to the results of statistical analyses of the data, detailed in Section 8.1.

To visualise errors generated by both the Bootstrapping and error ensemble methods, the BD of the M soil horizon TEST 5 was varied in increments of 0.01 in the range [0, 2], and the ensemble UCO_Ens_1 and error ensemble UCO_Err_Ens_9 re-tested. The standard deviation of the individual ANN predictions and the error ensemble predictions were calculated, and the results plotted as $\theta_{FC}$ *versus* BD, Figure 8.4a, allowing the direct comparison of the error ensemble and Bootstrapping methods of error prediction.

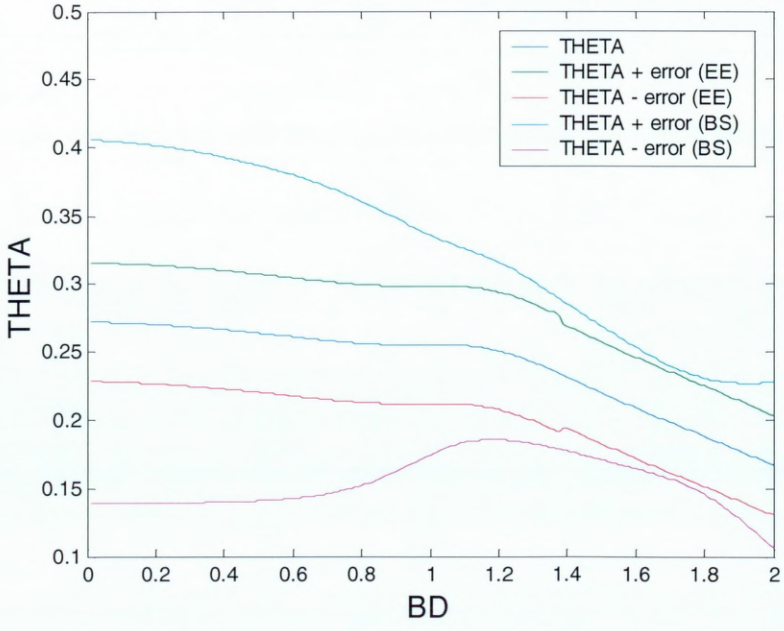



*Figure 8.4a: Effect of varying BD on M soil horizon TEST 5 on ensemble UCO_Ens_1, with error ensemble UCO_Err_Ens_9 (EE) and Bootstrapping (BS) error bars*

Generally, as BD increases, the ensemble prediction of $\theta_{FC}$ decreases linearly, although the magnitude of the decrease is more pronounced in the range [1, 2] than in the range [0, 1]. As BD increases in the range [0, 2], the error ensemble errors are small and constant, indicating that the individual ANNs are in relative agreement as to what $\theta_{FC}$ is for this soil horizon, and confidence is high. Recall from Tables 8.1d and 8.1e, and Figure 8.2e, that the effective range of BD for M soils is [0.16, 2.01], with a mean of 1.31 (median of 1.35). Thus, it is expected that error bars in this range should be small, indicating that there is a high degree of confidence in the ensemble predictions, and indeed this is the case. Indeed, the error bars are constant throughout the entire range tested, indicating that confidence is high for all predictions of $\theta_{FC}$ when BD is in the range [0, 2]. The Bootstrapping method produces error bars that are wide in the regions [0, 1.2] and [1.8, 2], indicating a low degree of confidence in these regions, and similar width to the error ensemble method in the region [1.2, 1.8].

This method of visualising has also been applied to the variation of %OM for TEST 5 in increments of 0.1 in the range [0, 10]. Again, UCO_Ens_1 and UCO_Err_Ens_9 were re-tested for variations in %OM, and the standard deviation of the individual ANN predictions and error ensemble predictions calculated. Results are plotted in Figure 8.4b.

Generally, as %OM increases, the ensemble prediction of $\theta_{FC}$ increases. The error ensemble method of error prediction produces error bars that are small and constant. Recall from Tables 8.1d and 8.1e, and Figure 8.2f, that the range of %OM for M soils is [0.02, 9.97], with a mean of 1.88 (and a median of 1.35), and the majority of measured $\theta_{FC}$ values for M soil horizons fall in the range [0, 3]. Thus, it is expected that there should be high confidence in the ensemble predictions in the region [0, 3] where we have the most information about %OM. Narrow error ensemble error bars in this region confirm that confidence is high. Indeed, confidence is high throughout the entire tested range of %OM values for the error ensemble method. The Bootstrapping method produces error bars that are the same as their corresponding error ensemble predictions, and diverge considerably when %OM > 3.

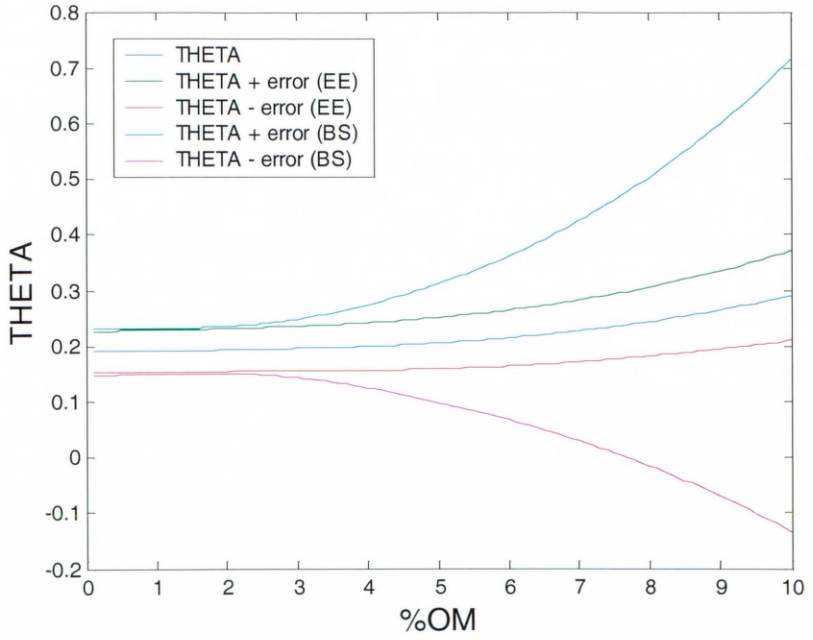



*Figure 8.4b: Effect of varying %OM on M soil horizon TEST 5 on ensemble UCO_Ens_1, with error ensemble UCO_Err_Ens_9 (EE) and Bootstrapping (BS) error bars*

Plotting $\theta_{FC}$ *versus* BD and %OM as a 3D surface plot allows the reader to see how $\theta_{FC}$ varies for both of these parameters. To do this, SSC of TEST 5 are held constant, whilst both BD and %OM are varied within their respective training ranges for model UCO_Ens_1, in increments of 0.02 and 0.1, respectively. The results are shown in the 3D surface plot of Figure 8.4c.

These results show that, for this particular test soil sample, $\theta_{FC}$ is a smoothly decreasing and increasing function of BD and %OM, respectively, with a saddle point running from high BD to high %OM. The ensemble model UCO_Ens_1 does not make any anomalous predictions of $\theta_{FC}$ on varying BD and %OM for this soil horizon and, since none are expected, this is a good sign. Recall that the model has little pedologic meaning for soils having BD < 0.16 (since the model would be extrapolating), and is of no use whatsoever when BD is equal to zero.

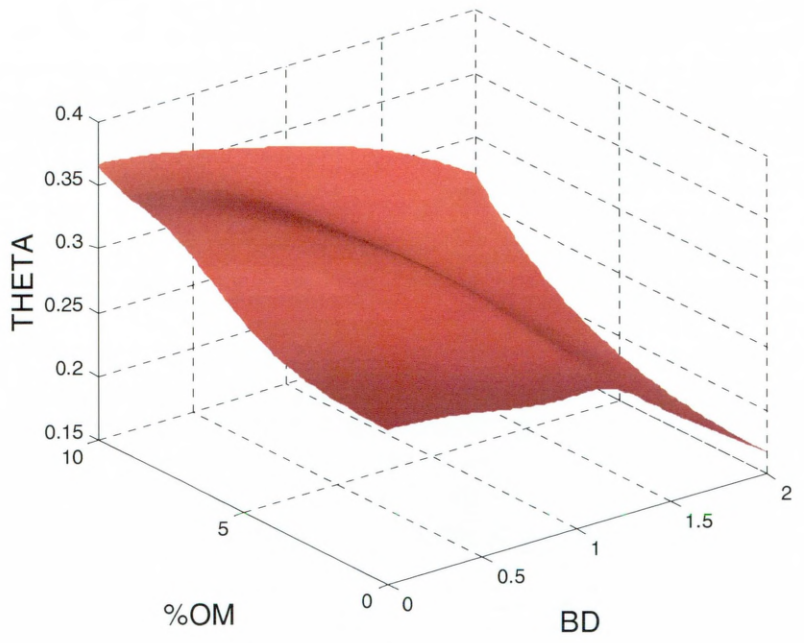



*Figure 8.4c: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble model UCO_Ens_1*

It is interesting to note that, according to Table 8.1e, on average, as %OM increases, $\theta_{FC}$ increases, and also that, on the average, as BD increases, $\theta_{FC}$ decreases, and this is exactly what we see in Figure 8.4c. In addition, the CV of %OM is high, Table 8.1g, indicating that %OM is highly variable, and perhaps more difficult to model accurately than parameters that are more consistent.

For the same M soil horizon, TEST 5, the error ensemble UCO_Err_Ens_9 may be tested in the same manner to determine which regions of this prediction surface are predicted with confidence (low error), and which areas should be treated with caution. The results of this are shown in the 3D surface plot of Figure 8.4d, which plots error *versus* BD and %OM.

Figure 8.4d shows that there is a smooth relationship between prediction error, BD and %OM. Note that the errors are greatest when %OM is large, and also that there is a 'step' anomaly running from %OM = 0 (BD = 1.5) to %OM = 10 (BD = 1). Since the range of %OM for M soils (Tables 8.1d and 8.1e, and Figure 8.2f) is [0.02, 9.97], with a mean of 1.88 (and a median of 1.35), this indicates that the distribution for %OM is strongly right-skewed (Section 8.1.5). Thus, it is expected that confidence should be highest when %OM is small, and this is indeed what can be seen in Figure

8.4d. In addition, since the effective range of BD for M soils (Tables 8.1d and 8.1e, and Figure 8.2f) is [0.16, 2.01], with a mean of 1.31 (median of 1.35), this indicates that the distribution of BD is left-skewed (Section 8.1.5). Therefore, confidence in the prediction of $\theta_{FC}$ should be highest when BD is large. Inspecting Figure 8.4d it can be seen that confidence is highest when BD is high and %OM is low, and this matches well to what should be expected.

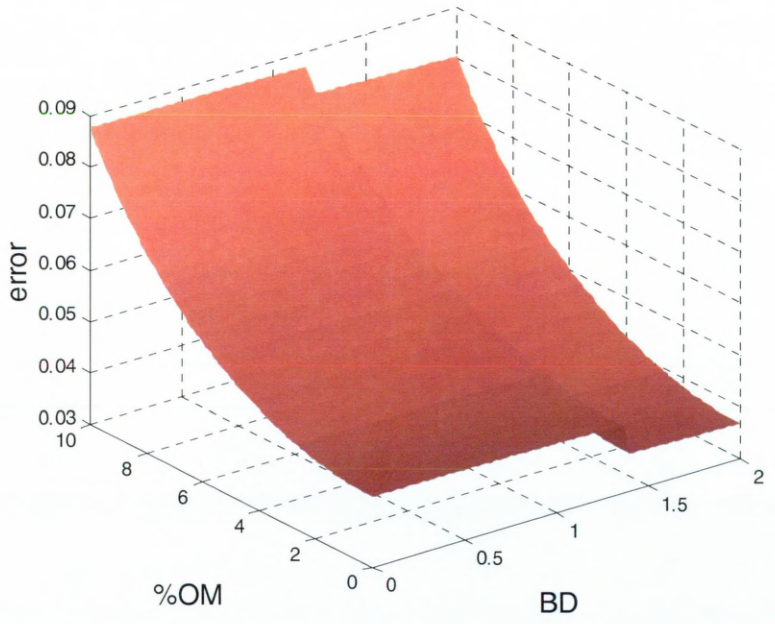



*Figure 8.4d: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble UCO_Ens_1 and error ensemble UCO_Err_Ens_9*

In order to assess how well this error surface matches with the error surface that would be predicted by the Bootstrapping method, the same procedure has been carried out on TEST 5 and the results plotted as Figure 8.4e.

The Bootstrapping method of error prediction shows that there is a smooth relationship between prediction error, BD and %OM. Note that the region of smallest error (and therefore highest confidence) is when BD is large and %OM is small, corresponding to the effective ranges of [0.16, 2.01] and [0, 3] for BD and %OM, respectively.

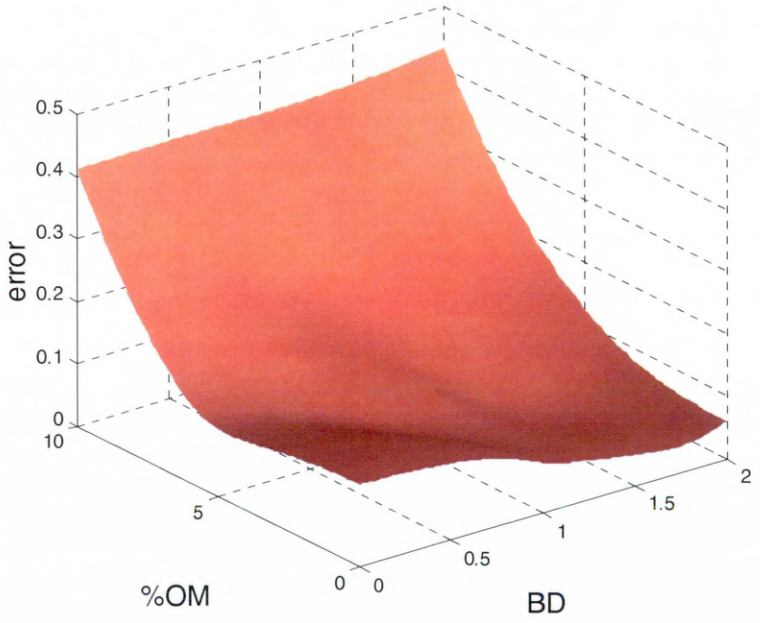



*Figure 8.4e: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble UCO_Ens_1 and the Bootstrapping method of error prediction*

### 8.4.2 Class PTF Error Bar Results

Analyses of the USDA-NRCS class PTF, error ensemble and Bootstrapping methods will be conducted in this section using M soil horizon TEST 5 (Table 6.2h), and results shown graphically as error bars and 3D error surfaces. These will enable a practical comparison of the two methods of error prediction, and will illustrate their comparative strengths and weaknesses. Additionally, the results of Section 8.3.1 will be discussed with reference to the results of statistical analyses of the data, detailed in Section 8.1.

To visualise errors generated by both the Bootstrapping and error ensemble methods, the BD of the M soil horizon TEST 5 was varied in increments of 0.01 in the range [0, 2], and the ensemble UCL_Ens_5 and error ensemble UCL_Err_Ens_10 re-tested. The standard deviation of the individual ANN predictions and the error ensemble predictions were calculated, and the results plotted as $\theta_{FC}$ *versus* BD, Figure 8.4f, allowing the direct comparison of the error ensemble and Bootstrapping methods of error prediction.

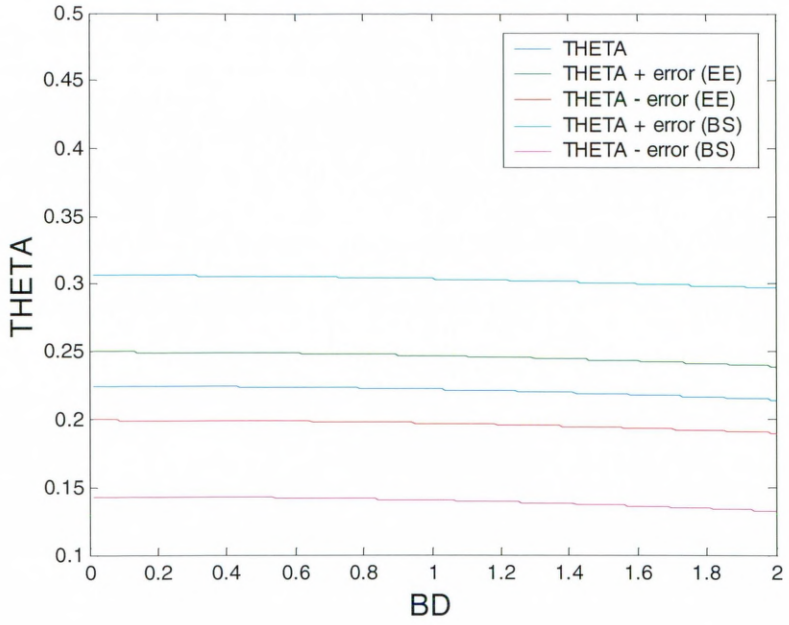



*Figure 8.4f: Effect of varying BD on M soil horizon TEST 5 on ensemble UCL_Ens_5, with error ensemble UCL_Err_Ens_10 (EE) and Bootstrapping (BS) error bars*

Generally, as BD increases, the ensemble prediction of $\theta_{FC}$ decreases linearly, although the magnitude of the decrease is small. As BD increases in the range [0, 2], the error ensemble errors are small and constant, indicating that the individual ANNs are in relative agreement as to what $\theta_{FC}$ is for this soil horizon, and confidence is high. Recall from Tables 8.1d and 8.1e, and Figure 8.2e, that the effective range of BD for M soils is [0.16, 2.01], with a mean of 1.31 (median of 1.35). Thus, it is expected that error bars in this range should be small, indicating that there is a high degree of confidence in the ensemble predictions, and indeed this is the case. Indeed, the error bars are constant throughout the entire range tested, indicating that confidence is high for all predictions of $\theta_{FC}$ when BD is in the range [0, 2]. The Bootstrapping method produces error bars that are wider than the corresponding error ensemble predictions, but are constant in the range [0, 2], indicating that confidence is equally high in all regions of BD.

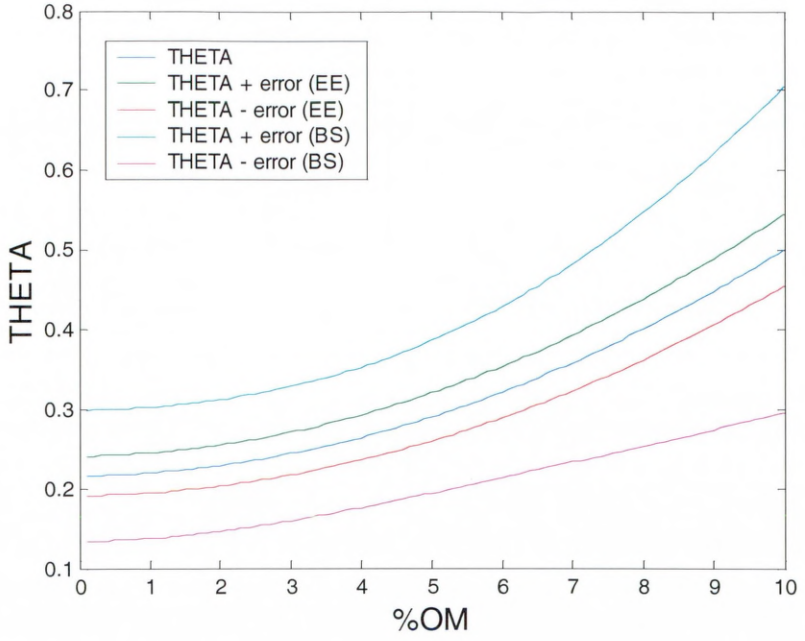



*Figure 8.4g: Effect of varying %OM on M soil horizon TEST 5 on ensemble UCL_Ens_5, with error ensemble UCL_Err_Ens_10 (EE) and Bootstrapping (BS) error bars*

This method of visualising has also been applied to the variation of %OM for TEST 5 in increments of 0.1 in the range [0, 10]. Again, UCL_Ens_5 and UCL_Err_Ens_10 were re-tested for variations in %OM, and the standard deviation of the individual ANN predictions and error ensemble predictions calculated. Results are plotted in Figure 8.4g.

Generally, as %OM increases, the ensemble prediction of $\theta_{FC}$ increases. The error ensemble method of error prediction produces error bars that are small and constant. Recall from Tables 8.1d and 8.1e, and Figure 8.2f, that the range of %OM for M soils is [0.02, 9.97], with a mean of 1.88 (and a median of 1.35), and the majority of measured $\theta_{FC}$ values for M soil horizons fall in the range [0, 3]. Thus, it is expected that there should be high confidence in the ensemble predictions in the region [0, 3] where we have the most information about %OM. Narrow error ensemble error bars in this region confirm that confidence is high. Indeed, confidence is high throughout the entire tested range of %OM values for the error ensemble method. The Bootstrapping method produces error bars that are wider than the corresponding error ensemble predictions, and diverge when %OM > 5.

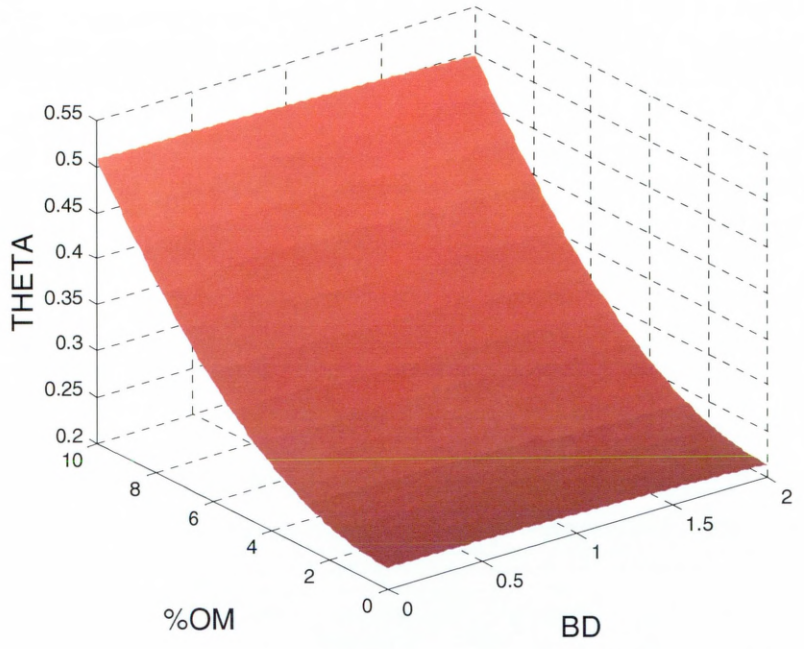



*Figure 8.4h: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble model UCL_Ens_5*

Plotting $\theta_{FC}$ *versus* BD and %OM as a 3D surface plot allows the reader to see how $\theta_{FC}$ varies for both of these parameters. To do this, SSC of TEST 5 are held constant, whilst both BD and %OM are varied within their respective training ranges for model UCL_Ens_5, in increments of 0.02 and 0.1, respectively. The results are shown in the 3D surface plot of Figure 8.4h.

These results show that, for this particular test soil sample, $\theta_{FC}$ is a smoothly decreasing and increasing function of BD and %OM, respectively. The ensemble model UCL_Ens_5 does not make any anomalous predictions of $\theta_{FC}$ on varying BD and %OM for this soil horizon. Recall that the model has little pedologic meaning for soils having BD < 0.16 (since the model would be extrapolating), and is of no use whatsoever when BD is equal to zero.

It is interesting to note that, according to Table 8.1e, on the average, as %OM increases, $\theta_{FC}$ increases, and also that, on the average, as BD increases, $\theta_{FC}$ decreases, and this is exactly what we see in Figure 8.4h.

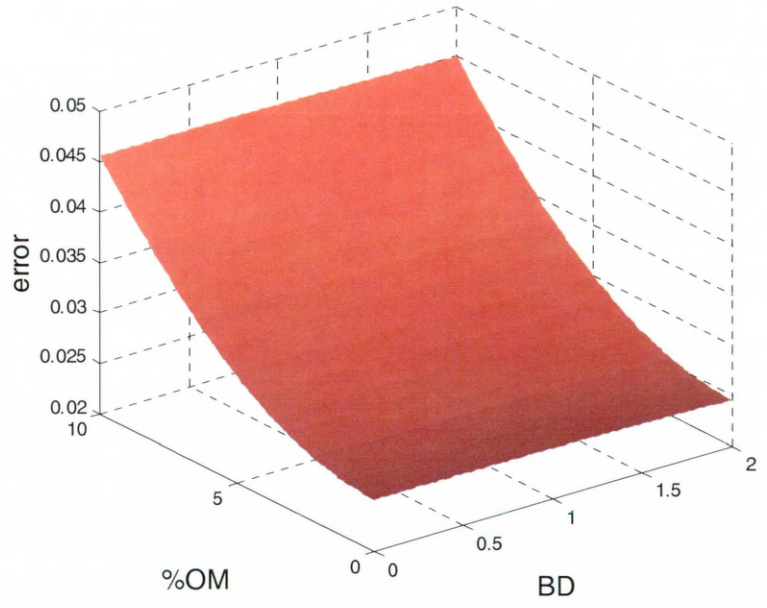



*Figure 8.4i: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble UCL_Ens_5 and error ensemble UCL_Err_Ens_10*

For the same M soil horizon, TEST 5, the error ensemble UCL_Err_Ens_10 may be tested in the same manner to determine which regions of this prediction surface are predicted with confidence (low error), and which areas should be treated with caution. The results of this are shown in the 3D surface plot of Figure 8.4i, which plots error *versus* BD and %OM.

Figure 8.4i shows that there is a smooth relationship between prediction error, BD and %OM. Note that the errors are greatest when %OM is large. Since the range of %OM for M soils (Tables 8.1d and 8.1e, and Figure 8.2f) is [0.02, 9.97], with a mean of 1.88 (and a median of 1.35), this indicates that the distribution for %OM is strongly right-skewed (Section 8.1.5). Thus, it is expected that confidence should be highest when %OM is small, and this is indeed what can be seen in Figure 8.4d. In addition, since the effective range of BD for M soils (Tables 8.1d and 8.1e, and Figure 8.2f) is [0.16, 2.01], with a mean of 1.31 (median of 1.35), this indicates that the distribution of BD is left-skewed (Section 8.1.5). Therefore, confidence in the prediction of $\theta_{FC}$ should be highest when BD is large. Inspecting Figure 8.4i it can be seen that confidence is unaffected by changes in BD.

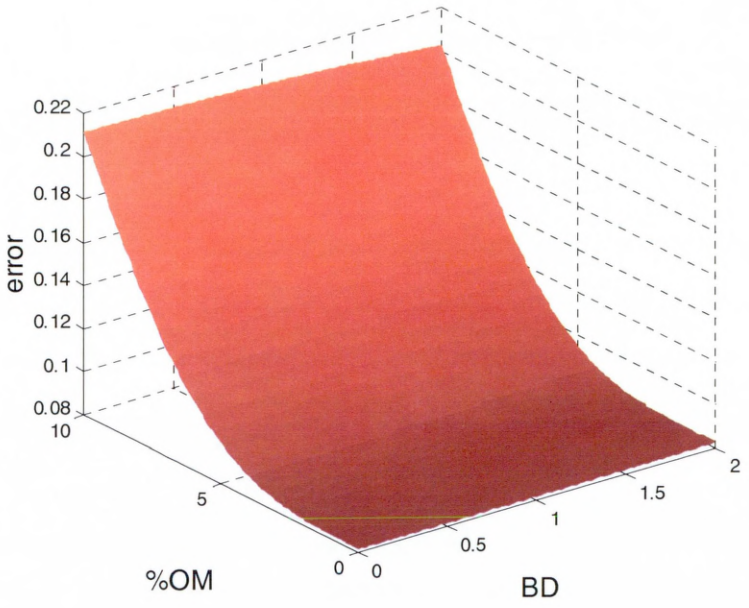



*Figure 8.4j: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble UCL_Ens_5 and the Bootstrapping method of error prediction*

In order to assess how well this error surface matches with the error surface that would be predicted by the Bootstrapping method, the same procedure has been carried out on TEST 5 and the results plotted as Figure 8.4j.

The Bootstrapping method of error prediction shows that there is a smooth relationship between prediction error, BD and %OM. Note that the region of smallest error (and therefore highest confidence) is when %OM is small, corresponding to the effective ranges of [0, 3] for %OM.

## Chapter Conclusions

Methods for constructing class and continuous PTFs of $\theta_{FC}$ by the ANN ensemble method have been discussed in detail in Chapters 6 and 7, and implemented using HYPRES data. Here, precisely the same methods have been used to construct class and continuous PTFs of $\theta_{FC}$ by the ANN ensemble method using data from the USDA-NRCS database. Although there is considerably more data available in this database than in HYPRES, the same selection criteria have been employed. Test results show that both class and continuous PTFs are stable and unbiased, and are

insensitive to the loss of any individual input parameter when sensitivity analyses are performed. Omitting BD & %OM results in a decrease in the ability of the continuous PTF to predict $\theta_{FC}$. Both class and continuous PTFs that have no soil textural component (i.e. only BD and %OM are present) have significantly decreased predictive ability, particularly with respect to the continuous PTF, where there is practically a complete loss of predictive ability.

The two competing methods of producing error bars on individual predictions (Bootstrapping and error ensemble) have been employed on both class and continuous PTFs, and found to produce similarly confident results in the region of high BD and low %OM, as expected.

# Chapter 9

# Model Comparisons

Chapters 6 to 8 detail the various models that feature in this thesis. Chapter 9 pools together all the tests, results and analyses of those models and direct comparisons are made here. The HYPRES class and continuous PTFs are compared, as are the USDA-NRCS class and continuous PTFs. In addition, the class PTFs for the HYPRES and USDA-NRCS databases are compared, and the same comparisons are made with the continuous PTFs. The class and continuous error models are also compared.

## 9.1 HYPRES Class and Continuous PTFs

Table 9.1a is a composite table, pooling together the continuous and class ANN-PTF RMSE results of Tables 6.2g and 7.2c, respectively. Additionally, Table 9.1a is plotted in Figure 9.1a.

There are no significant differences between the class and continuous PTFs when all input parameters are used, and also for single parameter omissions.

The class PTF yields improved predictions over the continuous PTF when SSC are omitted.

| Input parameters used | | | | | RMSE | |
|---|---|---|---|---|---|---|
| % Cl | % Si | % Sa | BD | %OM | M (continuous) | M (class) |
| ✓ | ✓ | ✓ | ✓ | ✓ | 0.059 | 0.061 |
| | ✓ | ✓ | ✓ | ✓ | 0.059 | 0.060 |
| ✓ | | ✓ | ✓ | ✓ | 0.059 | 0.060 |
| ✓ | ✓ | | ✓ | ✓ | 0.059 | 0.059 |
| ✓ | ✓ | ✓ | | ✓ | 0.063 | 0.062 |
| ✓ | ✓ | ✓ | ✓ | | 0.060 | 0.061 |
| ✓ | ✓ | ✓ | | | 0.070 | 0.068 |
| | | | ✓ | ✓ | 0.080 | 0.071 |

*Table 9.1a: Comparison of HYPRES RMSE results of sensitivity analyses of the continuous and class ANN-PTF ensemble methods, averaged over 10 models*

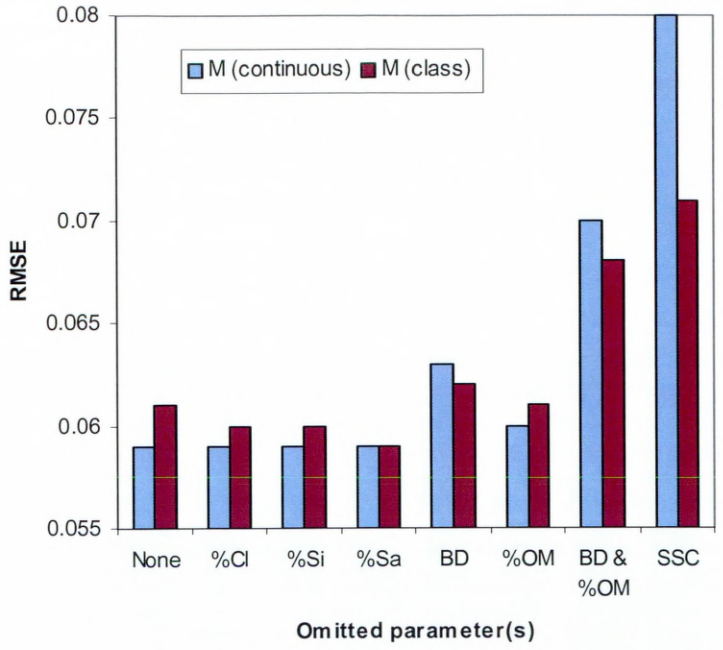



*Figure 9.1a: Comparison of HYPRES RMSE results of sensitivity analyses of the continuous and class ANN-PTF ensemble methods, averaged over 10 models*

Perhaps it is reasonable to expect that the class PTF should produce improved predictions of M soil horizons over the continuous PTF, since the class PTF is a specialised model, using more data in the training phase of the modelling process. Although this is not what the results of Table 9.1a and Figure 9.1a suggest, it may be inferred that the class PTF is overall more robust, since it is able to withstand the loss of multiple parameters better than the continuous PTF.

Figures 9.1b and 9.1c show how the ensemble-predicted $\theta_{FC}$ values for the M texture class correlate to the expected values, for the continuous PTF HCO_Ens_10 and the class PTF HCL_Ens_3, respectively.

It can be seen that the continuous and class PTFs predict $\theta_{FC}$ of M soil horizons equally well, since there are no significant differences between Figures 9.1b and 9.1c. This is in agreement with Table 9.1a and Figure 9.1a, where it was established that the $\theta_{FC}$ predictions of M soil horizons for the continuous and class PTFs (when all input parameters are used) are not significantly different.

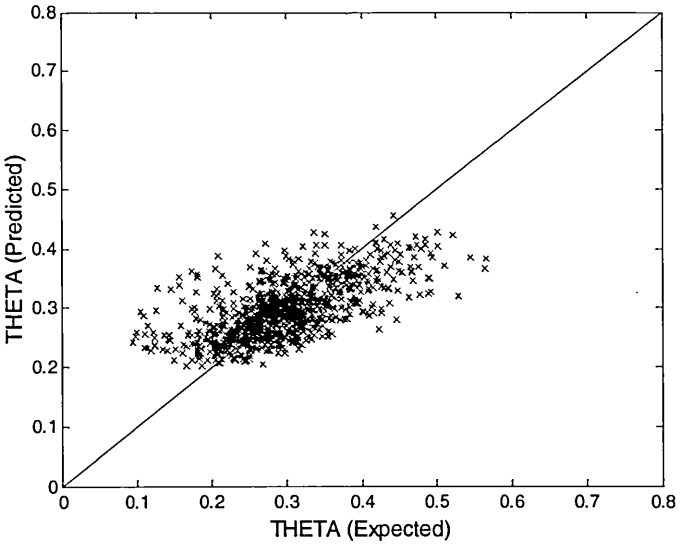



*Figure 9.1b: Comparison of expected and ensemble-predicted $\theta_{FC}$ for ensemble model HCO_Ens_10 of the M texture class*

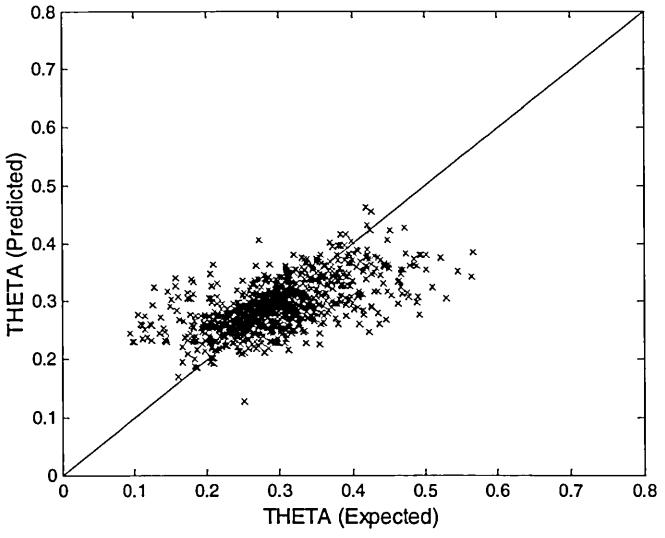



*Figure 9.1c: Comparison of expected and ensemble-predicted $\theta_{FC}$ for ensemble model HCL_Ens_3 of the M texture class*

When discussing evaluations of their regression-based class and continuous PTFs, Wösten *et al.* (1998) concluded that, although their class PTFs were of sufficient quality, '... *the predictions of the hydraulic characteristics when using continuous PTFs are not very accurate ... further analysis should reveal if the application of*

*neural networks instead of regression are feasible options.*' This supposition has been addressed in detail in this thesis. Utilising selected HYPRES data, it has been demonstrated that the ANN ensemble method of PTF construction yields class and continuous PTFs of comparable quality.

## 9.2 USDA-NRCS Class and Continuous PTFs

Table 9.2a is a composite table, pooling together the continuous and class ANN-PTF RMSE results of Tables 8.2d and 8.3c, respectively. Additionally, Table 9.2a is plotted in Figure 9.2a.

| Input parameters used | | | | | RMSE | |
|---|---|---|---|---|---|---|
| % Cl | % Si | % Sa | BD | %OM | M (continuous) | M (class) |
| ✓ | ✓ | ✓ | ✓ | ✓ | 0.065 | 0.066 |
| | ✓ | ✓ | ✓ | ✓ | 0.065 | 0.065 |
| ✓ | | ✓ | ✓ | ✓ | 0.065 | 0.066 |
| ✓ | ✓ | | ✓ | ✓ | 0.065 | 0.066 |
| ✓ | ✓ | ✓ | | ✓ | 0.065 | 0.065 |
| ✓ | ✓ | ✓ | ✓ | | 0.067 | 0.067 |
| ✓ | ✓ | ✓ | | | 0.070 | 0.069 |
| | | | ✓ | ✓ | *A* | 0.180 |

*Table 9.2a: Comparison of USDA-NRCS RMSE results of sensitivity analyses of the continuous and class ANN-PTF ensemble methods, averaged over 10 models, where A denotes that the results are detailed in Table 8.2e*

Overall, there are no differences between the predictive abilities of continuous and class PTFs of $\theta_{FC}$ for the M texture class, regardless of whether parameters are omitted, with the exception of the omission of SSC. It was established in Section 8.2.1 that depriving the USDA-NRCS continuous ANN-PTF ensemble of all textural information results in an unreliable model, and it may also be concluded that it has the same effect on the class PTF. Clearly, models based on the USDA-NRCS dataset are very sensitive to the removal of SSC.

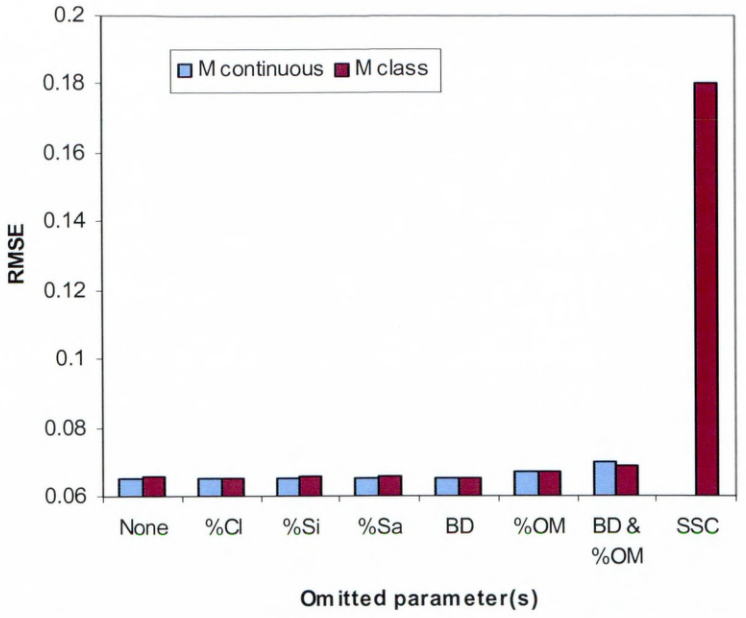



*Figure 9.2a: Comparison of USDA-NRCS RMSE results of sensitivity analyses of the continuous and class ANN-PTF ensemble methods, averaged over 10 models (the SSC omission of the continuous model is not shown)*

Figures 9.2b and 9.2c show how the ensemble-predicted $\theta_{FC}$ values for the M texture class correlate to the expected values, for the continuous PTF UCO_Ens_1 and the class PTF UCL_Ens_5, respectively.

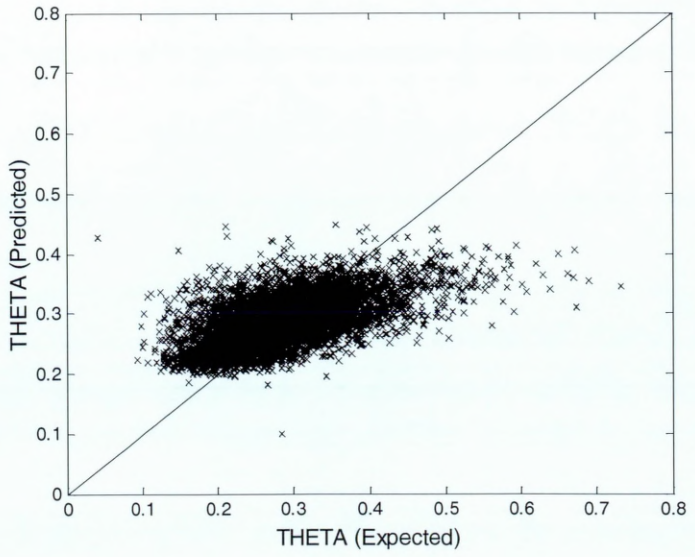



*Figure 9.2b: Comparison of expected and ensemble-predicted $\theta_{FC}$ for ensemble model UCO_Ens_1 of the M texture class*

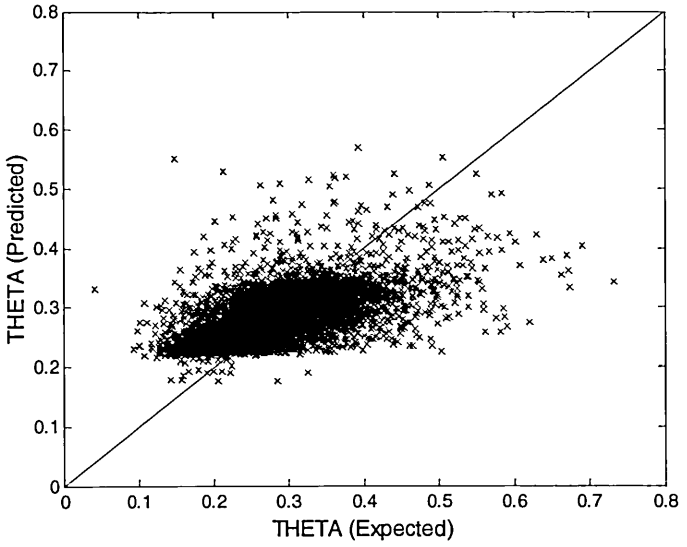



*Figure 9.2c: Comparison of expected and ensemble-predicted $\theta_{FC}$ for ensemble model UCL_Ens_5 of the M texture class*

It can be seen that the continuous and class PTFs predict $\theta_{FC}$ of M soil horizons equally well, since there are no significant differences between Figures 9.2b and 9.2c. This is in agreement with Table 9.2a and Figure 9.2a, where it was established that the $\theta_{FC}$ predictions of M soil horizons for the continuous and class PTFs (when all input parameters are used) are not significantly different.

Wösten *et al.* (1998) raised the supposition that ANNs may yield more accurate continuous PTFs than regression-based models. In Section 9.1 it was established that the ANN ensemble method of water retention PTF construction yields class and continuous PTFs of comparable quality, when HYPRES data are used. In parallel with this, it has been shown here that, using identical criteria and methods, class and continuous PTFs using USDA-NRCS data are also of comparable quality. This provides a confirmation that the similarities of the class and continuous PTF results are not emergent phenomena of the HYPRES dataset, but rather are the consequence of using the ensemble method. It may be concluded, therefore, that the ensemble method yields class and continuous PTFs of water retention that are of sufficient quality.

## 9.3 HYPRES and USDA-NRCS Continuous PTFs

Since the model used to construct a continuous ANN-PTF ensemble of water retention at field capacity remains unchanged for both the HYPRES and USDA-NRCS datasets, the results may be compared to assess the suitability of each for such a model. However, since the FC points for these two databases differ slightly, caution must be exercised and I wish to make it clear that the results of comparisons only indicate trends that require further investigations to be confirmed. Figure 9.3a is a composite plot of RMSE *versus* omitted parameter(s) for texture classes C to VF and overall (the RMSEs are extracted from Tables 6.2g and 8.2d). The plots of SSC as the omitted parameters for the USDA-NRCS dataset are not displayed on Figure 9.3a, since the ensemble results for these parameter omissions have been shown to be unreliable. Needless to say, the RMSEs for SSC parameter omissions for USDA-NRCS are considerably higher than their respective HYPRES results.

The predictive abilities of the ensemble method for the continuous PTF of $\theta_{FC}$ for the C texture class (Figure 9.3a) are better when data from HYPRES are used, irrespective of any parameter omissions. Both datasets reacted to the omission of parameters for C soils in the same manner, and to the same degree.

As with the C texture class, the ensemble model for M soils is more reliable when HYPRES data is used, irrespective of parameter omissions. The HYPRES dataset is more sensitive than USDA-NRCS with respect to the omission of BD from the ensemble model for M textured soils.

The ensemble model for MF soils is more reliable when HYPRES data is used, irrespective of parameter omissions, with the exception of BD & %OM. The HYPRES dataset is more sensitive than USDA-NRCS with respect to the omission of BD & %OM from the ensemble model for MF textured soils.

The predictive abilities of the ensemble method for the continuous PTF of $\theta_{FC}$ for the F texture class are comparable for both HYPRES and USDA-NRCS data, with the exception of the omission of BD and BD & %OM where HYPRES data is considerably more sensitive. For all parameter omissions (with the exception of SSC) the USDA-NRCS dataset is unaffected.

The ensemble model for VF soils is more reliable when HYPRES data is used, irrespective of parameter omissions. The HYPRES dataset is more sensitive than

USDA-NRCS with respect to the omission of BD and BD & %OM from the ensemble model for M textured soils.

Overall, the predictive abilities of the ensemble method for the continuous PTF of $\theta_{FC}$ are better when data from HYPRES is used, irrespective of any parameter omissions. The omission of any individual parameter has no effect on the overall predictive abilities of the ensemble, with the exception of the omission of BD from the HYPRES-trained ensemble. In addition, the model is adversely affected by the omission of BD & %OM and SSC, regardless of the dataset used.

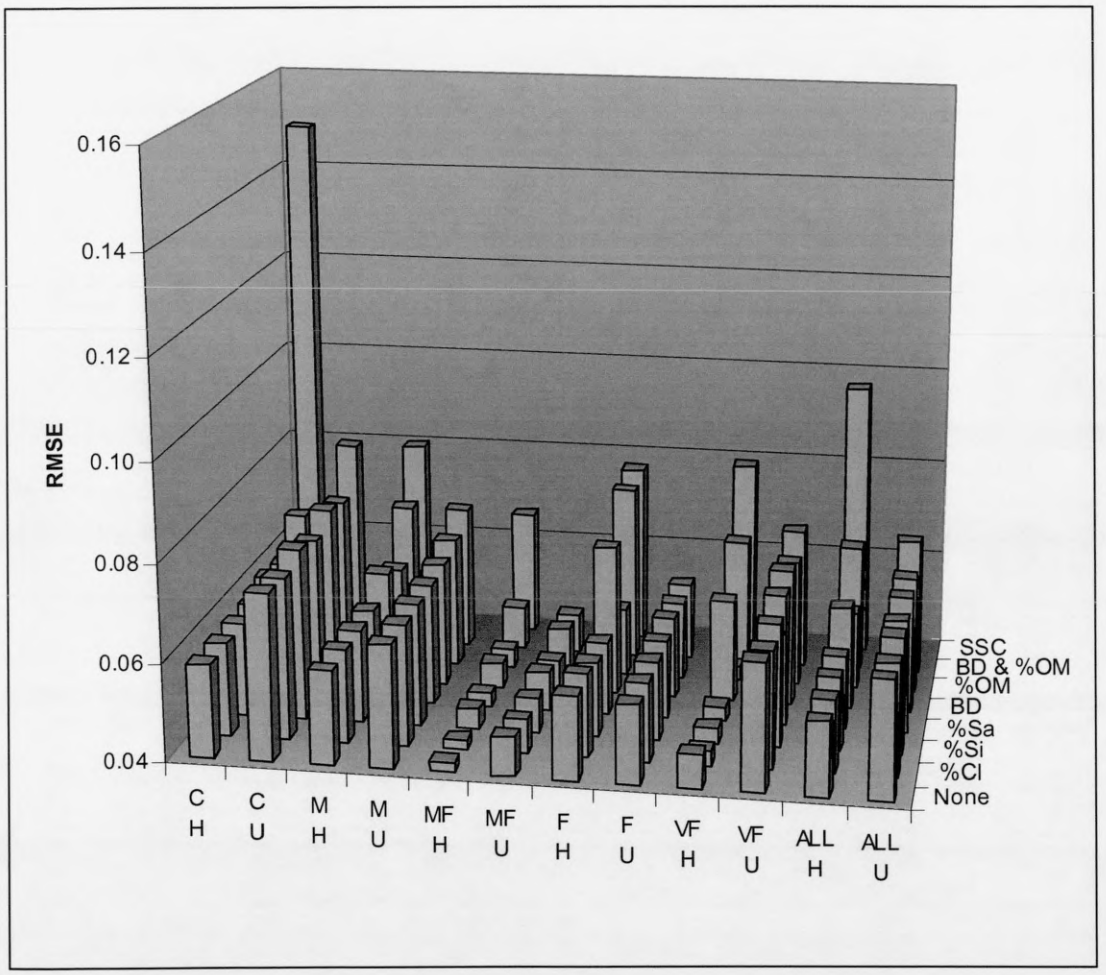



*Figure 9.3a: Comparison of RMSE results of sensitivity analyses for HYPRES and USDA-NRCS datasets. On the x-axis, H and U indicate HYPRES and USDA-NRCS results, respectively. Ensemble omitting SSC results are not shown for USDA-NRCS*

Figures 9.3b and 9.3c plot the expected *versus* ensemble-predicted values of $\theta_{FC}$ for HYPRES (model HCO_Ens_10) and USDA-NRCS (model UCO_Ens_1), respectively.

The expected *versus* ensemble-predicted $\theta_{FC}$ for the HYPRES dataset (Figure 9.3b) are better correlated with the expected 1:1 correspondence than the ensemble predictions from the USDA-NRCS dataset (Figure 9.3c) for all texture classes and overall.

In conclusion, the ensemble method produces a more reliable continuous PTF of $\theta_{FC}$ when trained and tested with data from HYPRES, as compared with data from USDA-NRCS. However, the results from both datasets used show that the ensemble model is a very effective method of constructing continuous PTFs of $\theta_{FC}$, requiring considerably less data than single ANN methods to produce the same degree of predictive ability.

As mentioned previously, since the FC points for these two databases differ slightly, caution must be exercised and the results of comparisons only indicate trends that require further investigations to be confirmed.

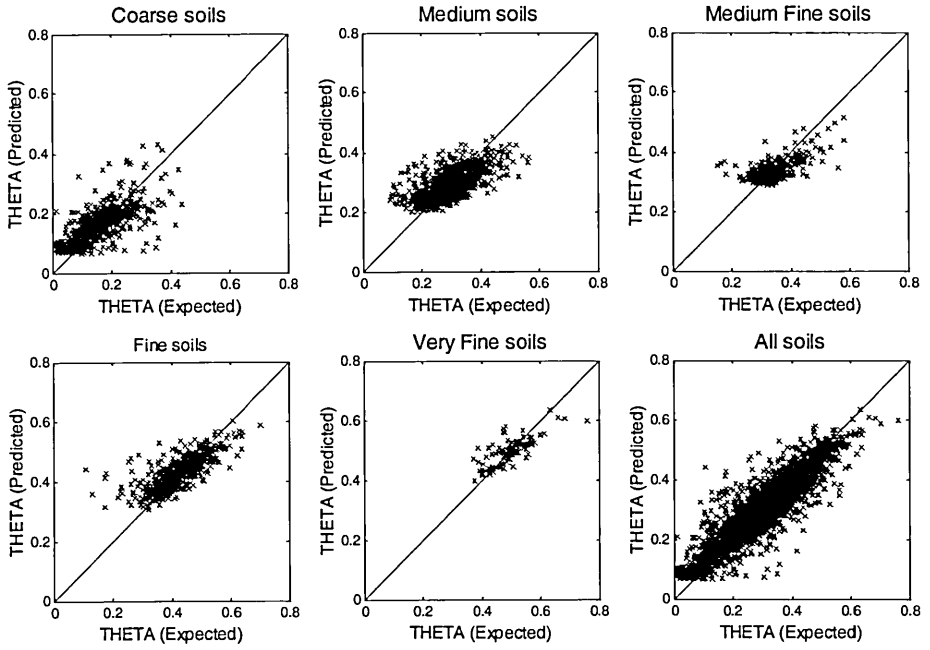



*Figure 9.3b: Comparison of expected and ensemble-predicted $\theta_{FC}$ for HYPRES ensemble model HCO_Ens_10, detailed by texture class and overall*

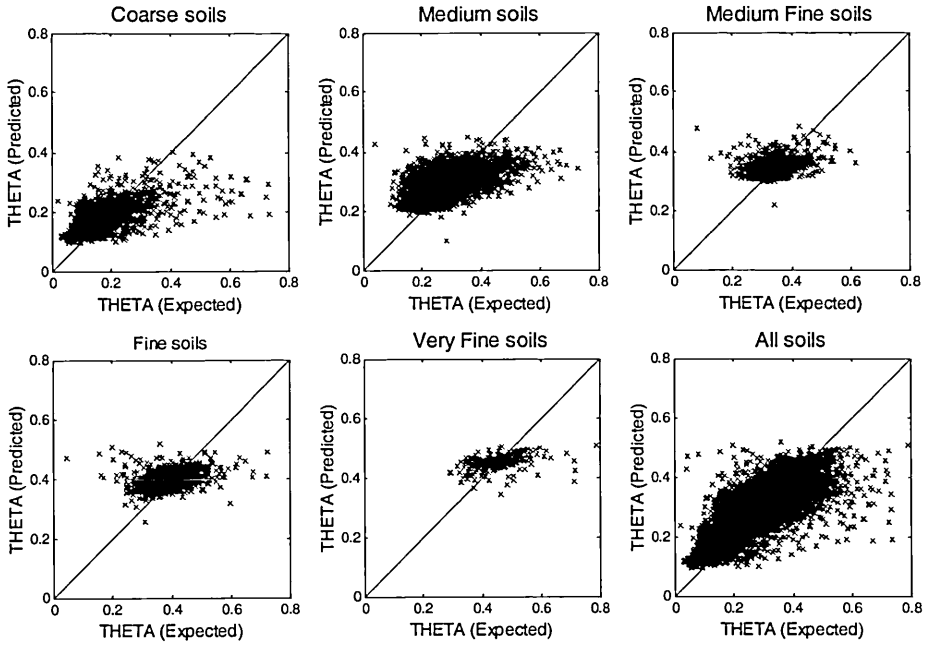



*Figure 9.3c: Comparison of expected and ensemble-predicted $\theta_{FC}$ for USDA-NRCS ensemble model UCO_Ens_1, detailed by texture class and overall*

## 9.4 HYPRES and USDA-NRCS Class PTFs

Since the model used to construct a class ANN-PTF ensemble of water retention at field capacity remains unchanged for both the HYPRES and USDA-NRCS datasets, the results may be compared to assess the suitability of each for such a model. Figure 9.4a is a plot of RMSE *versus* omitted parameter(s) for texture class M (the RMSEs are extracted from Tables 7.2c and 8.3c).

The ensemble model for M soils is more reliable when HYPRES data is used, irrespective of parameter omissions, but particularly so in the case of the omission of SSC. The HYPRES dataset is more sensitive than USDA-NRCS with respect to the omission of BD and BD & %OM from the class PTF ensemble model for M textured soils.

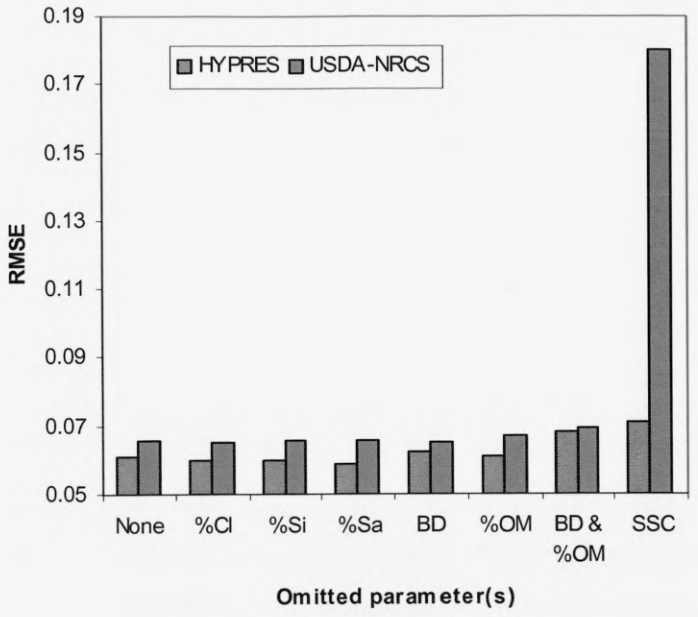



*Figure 9.4a: Comparison of M soil horizon RMSE results of sensitivity analyses for HYPRES and USDA-NRCS datasets*

Figures 9.4b and 9.4c plot the expected *versus* ensemble-predicted values of $\theta_{FC}$ for HYPRES (model HCL_Ens_3) and USDA-NRCS (model UCL_Ens_5), respectively.

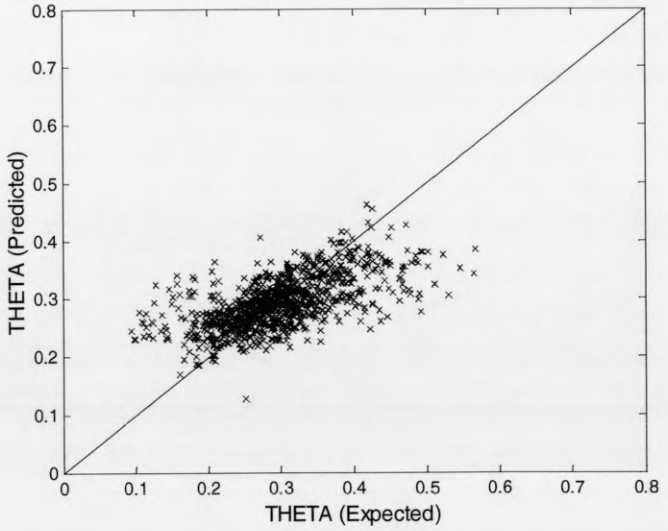



*Figure 9.4b: Comparison of expected and ensemble-predicted $\theta_{FC}$ for ensemble model HCL_Ens_3 of the M texture class*

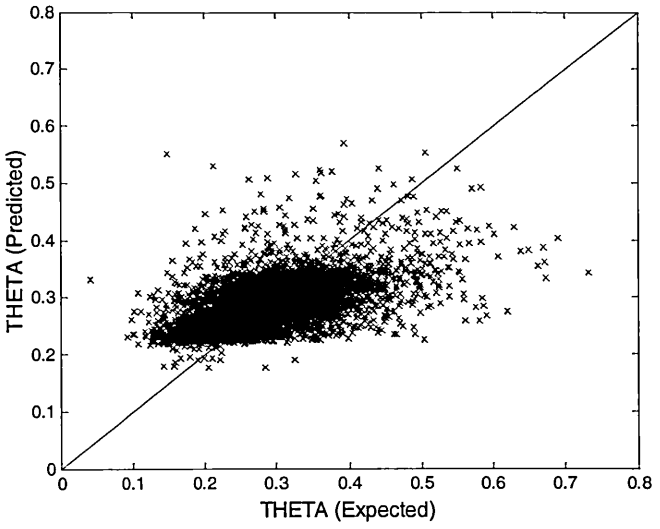



*Figure 9.4c: Comparison of expected and ensemble-predicted $\theta_{FC}$ for ensemble model UCL_Ens_5 of the M texture class*

Although it is noticeable that there are many more data points plotted for the USDA-NRCS dataset, there are no other significant differences between the predictive abilities of the HYPRES and USDA-NRCS class PTFs of $\theta_{FC}$ for M soil horizons.

## 9.5 HYPRES Class and Continuous Error Models

Comparisons of the HYPRES class and continuous PTFs, and the error ensemble and Bootstrapping error prediction methods will be conducted in this section, using M soil horizon TEST 5 (Table 6.2h). This will enable a practical comparison of the two competing methods of PTF construction, and the two competing methods of error prediction, and will illustrate their comparative strengths and weaknesses.

Figure 9.5a is a repeat of Figure 6.4k, and shows the variation of $\theta_{FC}$ throughout the range of possible values for BD and %OM, for the continuous PTF ensemble HCO_Ens_10, tested with M soil horizon TEST 5 (Table 6.2h). Also, Figure 9.5b is a repeat of Figure 7.3c, and shows the variation of $\theta_{FC}$ throughout the range of possible

values for BD and %OM, for the class PTF ensemble HCL_Ens_3, tested with M soil horizon TEST 5 (Table 6.2h).

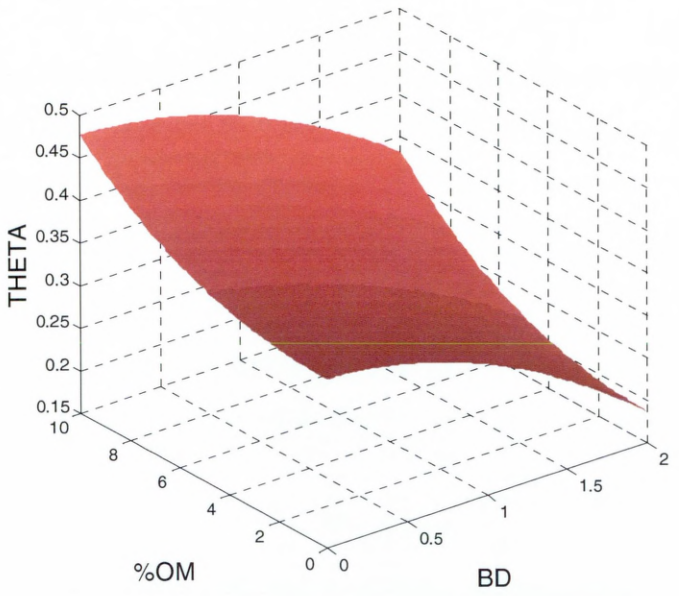



*Figure 9.5a: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble model HCO_Ens_10*

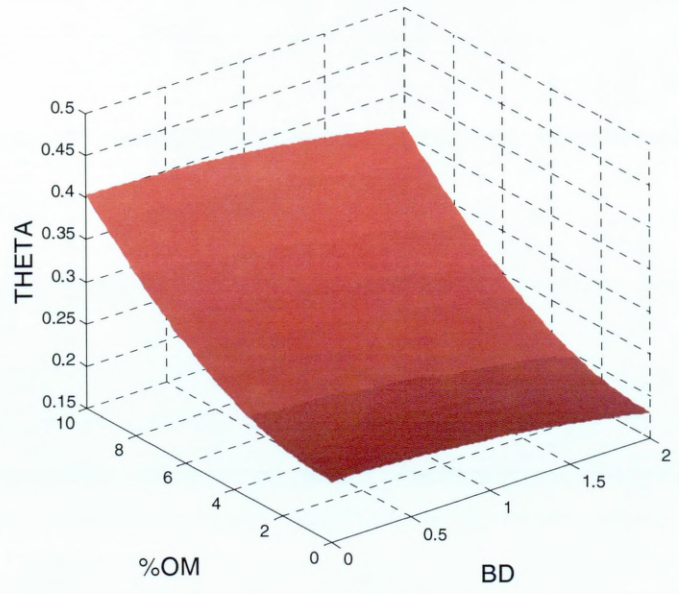



*Figure 9.5b: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble model HCL_Ens_3*

Both class and continuous PTFs show that $\theta_{FC}$ is a smoothly increasing and decreasing function of %OM and BD, respectively. The effect of increasing %OM is approximately equivalent in the two models for soil horizon TEST 5, whereas the effect on $\theta_{FC}$ of increasing BD is much greater for the continuous PTF than for the class PTF, regardless of the value of %OM. The region of highest confidence for both models is when BD is large and %OM is small, and the predictions of $\theta_{FC}$ in this region are approximately equal for both class and continuous PTFs.

Figure 9.5c is a repeat of Figure 6.4l, and shows the error predicted by the error ensemble method throughout the range of possible values for BD and %OM, for the continuous PTF ensemble HCO_Ens_10 and error ensemble HCO_Err_Ens_4, tested with M soil horizon TEST 5 (Table 6.2h). Also, Figure 9.5d is a repeat of Figure 7.3d, and shows the error predicted by the error ensemble method throughout the range of possible values for BD and %OM, for the class PTF ensemble HCL_Ens_3 and for error ensemble HCL_Err_Ens_1, tested with M soil horizon TEST 5 (Table 6.2h).

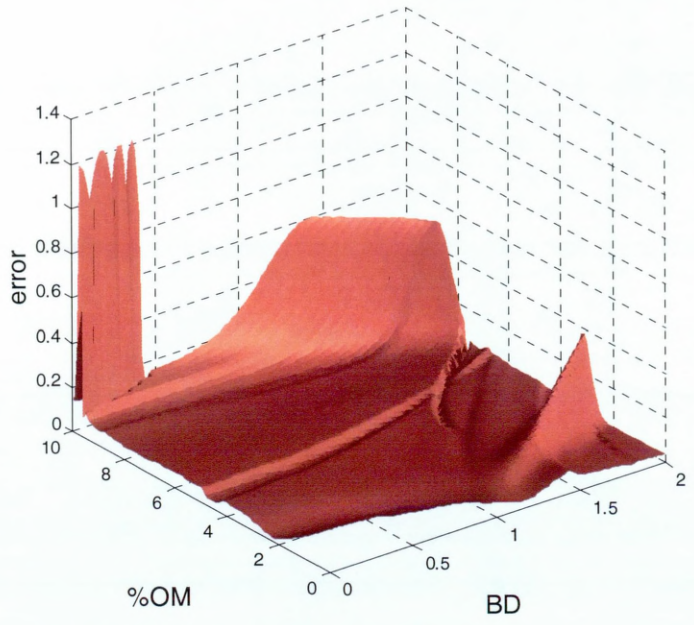



*Figure 9.5c: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble HCO_Ens_10 and error ensemble HCO_Err_Ens_4*

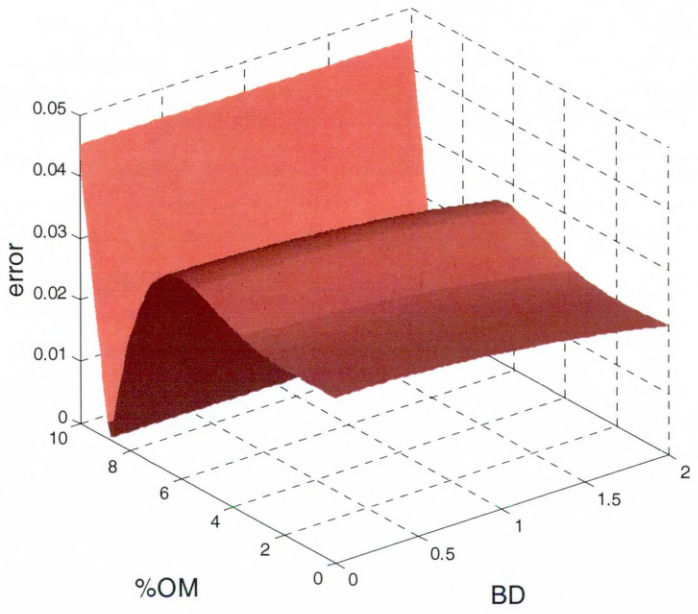



*Figure 9.5d: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble HCL_Ens_3 and for error ensemble HCL_Err_Ens_1*

Both class and continuous error ensemble methods have quite a flat error response for variations in BD and %OM, with the exception of the anomalous features, mostly on the continuous error ensemble plot. It is worth noting that most of the anomalous features on both the class and continuous error ensemble plots occur in regions of high %OM. Recall that prediction confidence is highest when %OM is low, specifically in the region [0, 3], thus, predictions in the higher regions of %OM are to be treated with caution. When %OM is low, the magnitude of the errors for the class and continuous PTFs are in the range [0.02, 0.05].

Figure 9.5e is a repeat of Figure 6.4n, and shows the error predicted by the Bootstrapping method throughout the range of possible values for BD and %OM, for the continuous PTF ensemble HCO_Ens_10, tested with M soil horizon TEST 5 (Table 6.2h). Also, Figure 9.5f is a repeat of Figure 7.3e, and shows the error predicted by the Bootstrapping method throughout the range of possible values for BD and %OM, for the class PTF ensemble HCL_Ens_3, tested with M soil horizon TEST 5 (Table 6.2h).

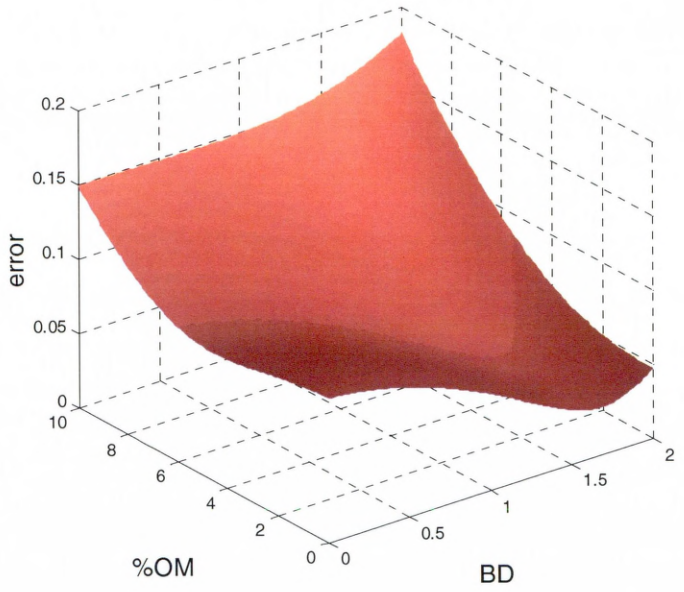



*Figure 9.5e: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble HCO_Ens_10 and the Bootstrapping method of error prediction*

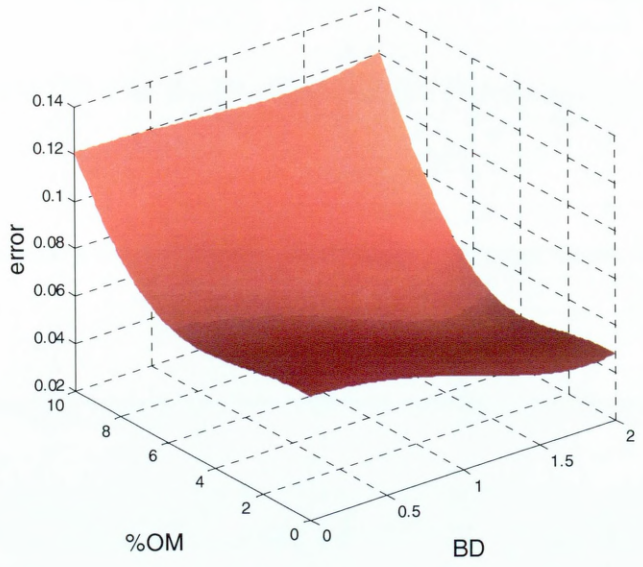



*Figure 9.5f: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble HCL_Ens_3 and for the Bootstrapping method of error prediction*

The 3D surface plots of Figures 9.5e and 9.5f show that the Bootstrapping method of error prediction produces similar results for both the class and continuous PTFs. The highest confidence is expected to be in the region where BD is high and %OM is low, and this is exactly what is seen. In this region, the magnitude of the predicted error is in the range [0.03, 0.05], and this matches well to the reliable range of errors in the error ensemble method, although the Bootstrapping method produces a much smoother error function.

Overall, it would seem that both error ensemble and Bootstrapping methods are valid techniques of producing error prediction, although this conclusion should be treated with caution, since the competing methods have only been investigated with a single test sample.

## 9.6 USDA-NRCS Class and Continuous Error Models

Comparisons of the USDA-NRCS class and continuous PTFs, and the error ensemble and Bootstrapping error prediction methods will be conducted in this section, using M soil horizon TEST 5 (Table 6.2h). This will enable a practical comparison of the two competing methods of PTF construction, and the two competing methods of error prediction, and will illustrate their comparative strengths and weaknesses.

Figure 9.6a is a repeat of Figure 8.4c, and shows the variation of $\theta_{FC}$ throughout the range of possible values for BD and %OM, for the continuous PTF ensemble UCO_Ens_1, tested with M soil horizon TEST 5 (Table 6.2h). Also, Figure 9.6b is a repeat of Figure 8.4h, and shows the variation of $\theta_{FC}$ throughout the range of possible values for BD and %OM, for the class PTF ensemble UCL_Ens_5, tested with M soil horizon TEST 5 (Table 6.2h).

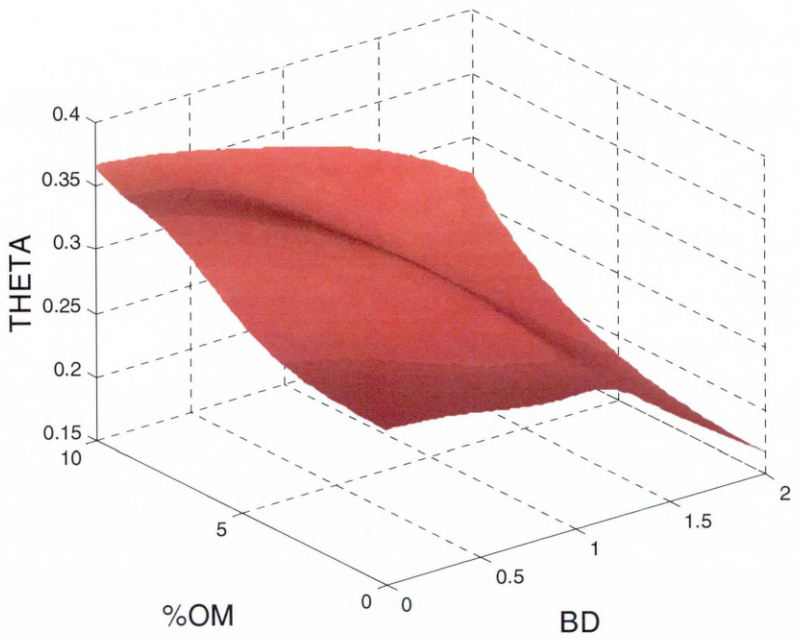



*Figure 9.6a: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble model UCO_Ens_1*

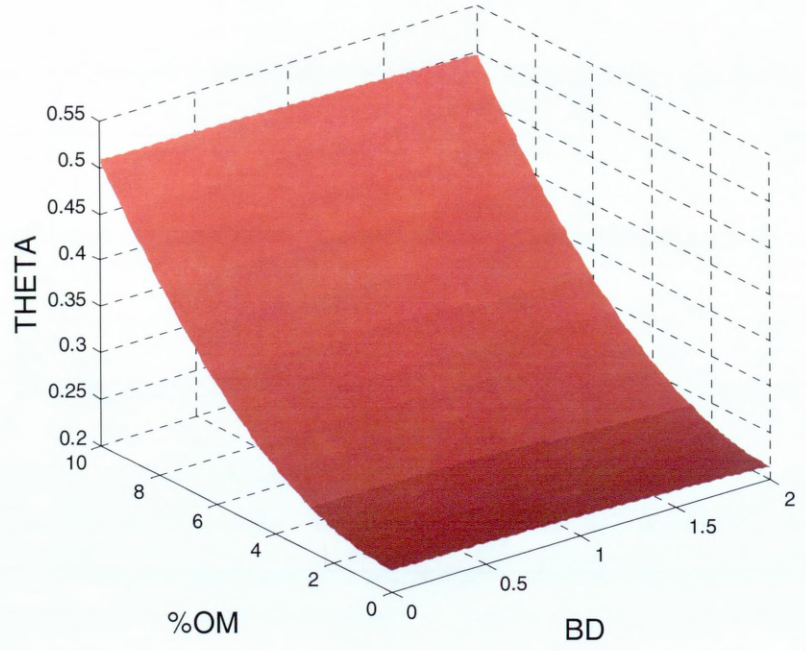



*Figure 9.6b: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble model UCL_Ens_5*

Both class and continuous PTFs show that $\theta_{FC}$ is a smoothly increasing and decreasing function of %OM and BD, respectively. The effect of increasing %OM is more pronounced in the class PTF for soil horizon TEST 5, whereas the effect on $\theta_{FC}$ of increasing BD is much greater for the continuous PTF than for the class PTF, regardless of the value of %OM. The region of highest confidence for both models is when BD is large and %OM is small. In this region predictions of $\theta_{FC}$ are smaller for the continuous PTF than for the class PTF.

Figure 9.6c is a repeat of Figure 8.4d, and shows the error predicted by the error ensemble method throughout the range of possible values for BD and %OM, for the continuous PTF ensemble UCO_Ens_1 and error ensemble UCO_Err_Ens_9, tested with M soil horizon TEST 5 (Table 6.2h). Also, Figure 9.6d is a repeat of Figure 8.4i, and shows the error predicted by the error ensemble method throughout the range of possible values for BD and %OM, for the class PTF ensemble UCL_Ens_5 and for error ensemble UCL_Err_Ens_10, tested with M soil horizon TEST 5 (Table 6.2h).

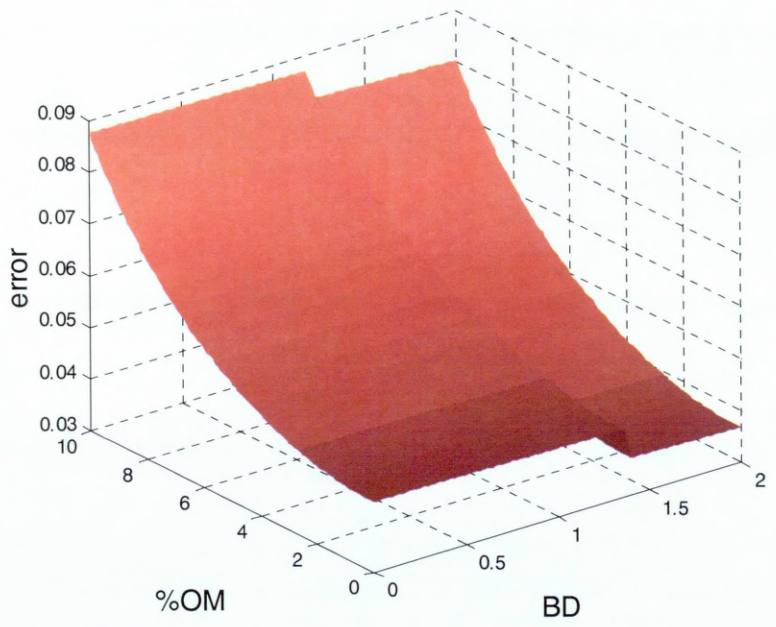



*Figure 9.6c: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble UCO_Ens_1 and error ensemble UCO_Err_Ens_9*

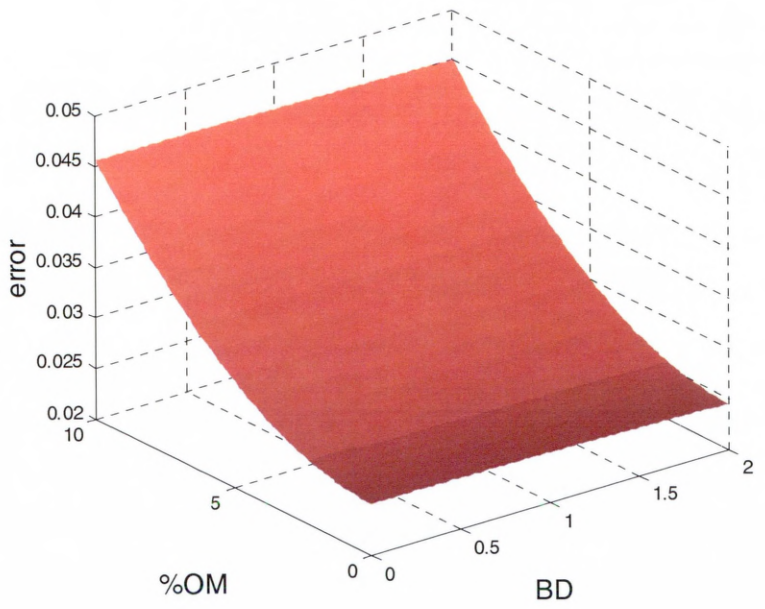



*Figure 9.6d: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble UCL_Ens_5 and error ensemble UCL_Err_Ens_10*

Both class and continuous error ensemble methods have an error response that increases as %OM is increased, but is affected little by changing BD. This reflects the fact that most information about %OM is in the range [0, 3], and thus, this is the range that is expected to produce the highest confidence in predictions. Overall, the continuous error ensemble produces errors that are larger than their corresponding class error ensemble predictions, regardless of the values of BD and %OM.

Figure 9.6e is a repeat of Figure 8.4e, and shows the error predicted by the Bootstrapping method throughout the range of possible values for BD and %OM, for the continuous PTF ensemble UCO_Ens_1, tested with M soil horizon TEST 5 (Table 6.2h). Also, Figure 9.6f is a repeat of Figure 8.4j, and shows the error predicted by the Bootstrapping method throughout the range of possible values for BD and %OM, for the class PTF ensemble UCL_Ens_5, tested with M soil horizon TEST 5 (Table 6.2h).

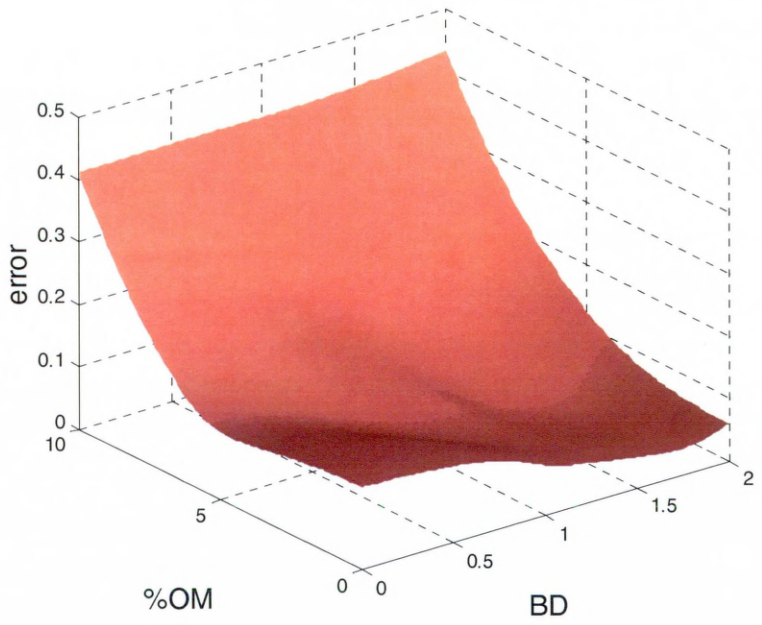



*Figure 9.6e: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble UCO_Ens_1 and the Bootstrapping method of error prediction*

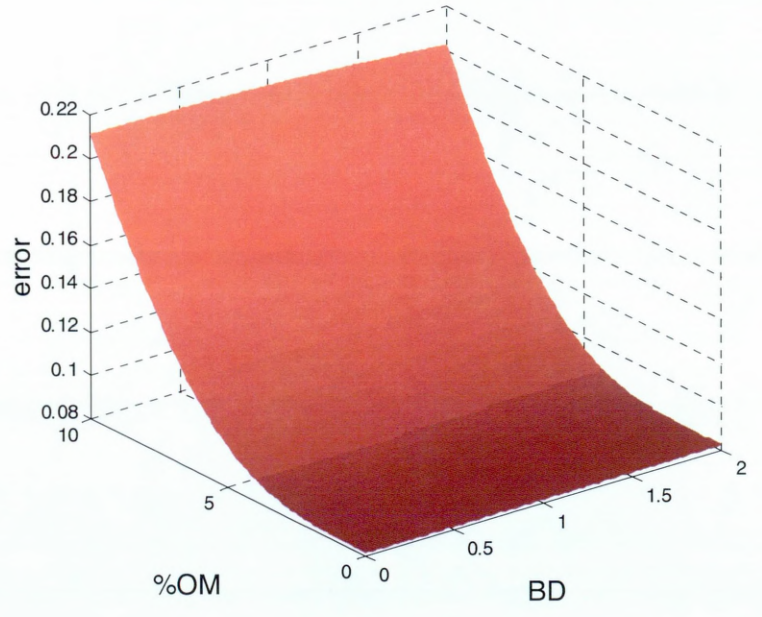



*Figure 9.6f: 3D surface plot of the results of varying BD and %OM of M soil horizon TEST 5, for ensemble UCL_Ens_5 and the Bootstrapping method of error prediction*

The 3D surface plots of Figures 9.6e and 9.6f show that the Bootstrapping method of error prediction produces similar surface plots for both the class and continuous PTFs, albeit that the results for the continuous PTF are typically twice as large as the corresponding values for the class PTF. The highest confidence is expected to be in the region where BD is high and %OM is low, and this is exactly what is seen for the continuous PTF, although errors for the class PTF appear to be independent of BD.

Overall, it would seem that both error ensemble and Bootstrapping methods are valid techniques of producing error prediction, although this conclusion should be treated with caution, since the competing methods have only been investigated with a single test sample.

## Chapter Conclusions

Since the methods for producing class and continuous PTFs are very similar, the results produced by each may be directly compared. Additionally, since the methods are consistent, the two differing databases may be compared to assess their relative suitabilities for constructing class and continuous PTFs.

For HYPRES, the continuous PTF produces more reliable predictions, but the class PTF is more robust in the sense that it is less sensitive to the loss of multiple input parameters from the model. For the USDA-NRCS database, the class and continuous PTFs produced virtually identical results, with the exception that, although the loss of SSC from the class PTF resulted in very poor results, the loss of SSC from the continuous PTF resulted in the (almost) complete collapse of the PTFs predictive ability.

For the continuous PTF, HYPRES data produced improved predictions over USDA-NRCS data for all texture classes, except for F soil horizons, where results were comparable. Exceptions to this are found when sensitivity analyses are performed, and these show that USDA-NRCS data performs better than HYPRES data in the MF texture class when BD & %OM are omitted, and in the F texture class when BD and BD & %OM are omitted.

For the class PTF, HYPRES data produced improved predictions over USDA-NRCS data for all texture classes regardless of any parameter omissions, although HYPRES is more sensitive to the omission of BD and BD & %OM.

All models, whether class or continuous, or HYPRES or USDA-NRCS showed that as BD increases $\theta_{FC}$ decreases, and that as %OM increases $\theta_{FC}$ increases.

For the HYPRES error ensemble, both class and continuous models have a flat error response, although both also have some anomalous features, particularly with respect to the continuous model.

The USDA-NRCS error ensemble class and continuous models both have smooth error functions, although the continuous model is more sensitive to changes in BD than the class model.

For the HYPRES Bootstrapping model, both class and continuous PTFs produce similar predictions of error, whilst for the USDA-NRCS Bootstrapping model, the continuous PTF produces error predictions that are twice as large as those predicted by the class PTF. In addition, the USDA-NRCS Bootstrapping method performed on the class PTF is independent of BD.

Regardless of which error prediction method is used, all produce similarly confident results in the region of high BD and low %OM, as expected.

# Chapter 10

# ANN-PTF Ensembles of Hydraulic Conductivity

It is one of the stated aims of this thesis that class and continuous artificial neural network ensemble pedotransfer functions of soil hydraulic conductivity at field capacity would be constructed using data from the HYPRES dataset. Significant amounts of work have been conducted in trying to achieve these specific aims, and results of attempts to construct such models are detailed and discussed here. These results will demonstrate why I will conclude that HYPRES does not contain sufficient data to allow the successful construction of class or continuous artificial neural network ensemble pedotransfer functions of hydraulic conductivity at field capacity.

## 10.1 Data Selection

Section 5.1 detailed the selection criteria applied to the data used in modelling water retention. These same criteria have been applied to soil horizons in HYPRES for hydraulic conductivity at field capacity, and the data that resulted are shown in Table 10.1a.

| | Texture Class | | | | | |
|---|---|---|---|---|---|---|
| | **C** | **M** | **MF** | **F** | **VF** | **ALL** |
| Total number of soil horizons selected | 155 | 138 | 157 | 69 | 19 | 538 |
| Number deleted where %OM > 10 | 0 | 0 | 0 | 0 | 0 | 0 |
| Number of deleted duplicates | 12 | 18 | 7 | 11 | 3 | 51 |
| **Total number remaining** | **143** | **120** | **150** | **58** | **16** | **487** |
| Percentage of total remaining | 29 | 25 | 31 | 12 | 3 | 100 |

*Table 10.1a: Textural breakdown of the data selected for modelling*

Of the 538 soil horizons selected prior to application of the selection criteria, 51 of these were deleted (all 51 were duplicates), leaving 487 viable soil horizons for the modelling process. Compare this figure with the amount of data available for modelling water retention – 2779 soil horizons.

Approximately one-quarter of the selected soil horizons belong to the M texture classes, whilst the C and MF texture classes contribute approximately one-third each. The F texture class contributes a little over one-tenth, whilst the remaining 3% belong to the little-represented VF texture class. The textural distribution of these soil horizons is shown on the soil textural triangle of Figure 10.1a.

Statistical analyses performed on these data will not be shown here.

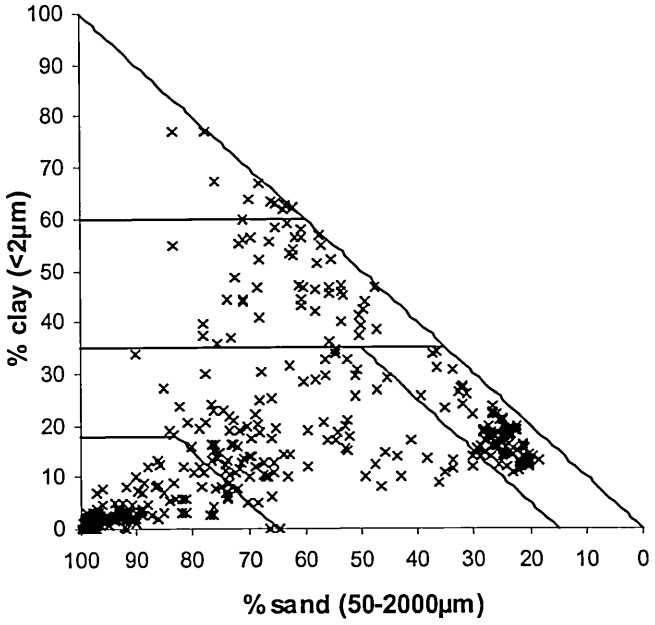



*Figure 10.1a: Textural distribution of the data chosen for the modelling process*

## 10.2 Modelling Procedure

The input parameters are, as with the modelling of soil water retention, proportions of sand, silt and clay, dry bulk density and organic matter content, and the output parameter is hydraulic conductivity at field capacity (-250cm$H_2O$). All input parameters are re-scaled by the algorithm in equation 6.1b. Water retention values are measured on a linear scale, and fall within the range [0, 1], however, hydraulic conductivity is measured logarithmically, and values fall in the range [$10^{-9}$, $10^{9}$]. Due to the nature of transfer functions, it is common practice to re-scale the input and output parameters of artificial neural networks, so that they fall, for instance, in the range [-1, 1] or [0, 1], depending on the transfer function used. Thus, the hydraulic conductivity values have been re-scaled by the following algorithm:

$$u_j = \frac{\log_{10}(output_j)}{10} \tag{10.2a}$$

This ensures that the output parameter values of the ANNs will fall within the range [-1, 1]. Using the tansig transfer function (recall that the logsig transfer function was used for water retention), training of the ANNs and construction of the ensemble may proceed in precisely the same way as in the water retention ensemble. Since there are only 16 VF soil horizons with which to construct a model, it was decided that a class ANN-PTF ensemble of hydraulic conductivity should be constructed in the first instance, using only MF data, since data in this texture class is most abundant.

## 10.3 Results

Table 10.3a details the results (RMSE and ME) of ten ensembles of hydraulic conductivity at field capacity, using 70 MF soil horizons in each of ten ensemble members.

| | MF | |
|---|---|---|
| | **RMSE** | **ME** |
| **K_Ens_1** | 0.068 | 0.0318 |
| **K_Ens_2** | 0.069 | 0.0321 |
| **K_Ens_3** | 0.068 | 0.0312 |
| **K_Ens_4** | 0.069 | 0.0321 |
| **K_Ens_5** | 0.069 | 0.0321 |
| **K_Ens_6** | 0.069 | 0.0321 |
| **K_Ens_7** | 0.069 | 0.0321 |
| **K_Ens_8** | 0.069 | 0.0321 |
| **K_Ens_9** | 0.069 | 0.0321 |
| **K_Ens_10** | 0.069 | 0.0320 |

*Table 10.3a: RMSEs and MEs of 10 ensembles*

On first inspection of the results, it appears that the RMSEs are very similar to the results for water retention (recall that the average RMSE = 0.061 for an M class ANN-PTF ensemble of $\theta_{FC}$). However, since the range of values for water retention and hydraulic conductivity are [0, 1] and [$10^{-9}$, $10^{9}$], respectively, the RMSE results

for such should not necessarily be of the same order of magnitude. To understand the degree of accuracy of the hydraulic conductivity results, it is necessary to inspect the results of individual tests. Table 10.3b shows the results of ten randomly chosen tests, and details the physical and hydraulic characteristics of each soil horizon, the ensemble prediction, the prediction error and the percentage error (%$\epsilon$).

The predictions of hydraulic conductivity are typically an order of magnitude higher or lower than the expected value, with the majority of predictions underestimating $K_{FC}$. Considering these ten individual test results, the overall RMSE is 0.044 (of course, on increasing the number of tests, the RMSE would approach the overall RMSE results of Table 10.3a). Clearly, predictions that are an order of magnitude different from their expected values cannot be referred to as 'accurate'. For an ANN-PTF ensemble model of hydraulic conductivity to be referred to as accurate, it must produce predictions that are at least of the same order of magnitude as the expected values, and this would then lead to an RMSE that is one order of magnitude smaller than those published in Tables 10.3a and 10.3b.

| %Cl | %Sa | %Si | BD | %OM | $K_{FC}$ | $\hat{K}_{FC}$ | $\hat{K}_{FC} - K_{FC}$ | %$\epsilon$ |
|---|---|---|---|---|---|---|---|---|
| 18.9 | 77.5 | 3.6 | 1.596 | 0 | 0.034 | 0.0015 | -0.0325 | -96 |
| 19.1 | 71.8 | 9.1 | 1.488 | 0.24 | 0.0035 | 0.0004 | -0.0031 | -89 |
| 13.7 | 79.4 | 6.9 | 1.317 | 2 | 0.059 | 0.0123 | -0.0467 | -79 |
| 11.1 | 77.9 | 11.0 | 1.400 | 0.78 | 0.054 | 0.0936 | 0.0396 | 73 |
| 16 | 74.2 | 9.8 | 1.31 | 1.68 | 0.034 | 0.0035 | -0.0305 | -90 |
| 13.3 | 79.4 | 7.3 | 1.48 | 1.9 | 0.097 | 0.1970 | 0.1000 | 103 |
| 14.8 | 71.9 | 13.3 | 1.48 | 1.72 | 0.016 | 0.0006 | -0.0154 | -96 |
| 22.2 | 74 | 3.8 | 1.521 | 0.2 | 0.016 | 0.0037 | -0.0123 | -77 |
| 13.3 | 79.4 | 7.3 | 1.334 | 1.9 | 0.071 | 0.0163 | -0.0547 | -77 |
| 18 | 73.6 | 8.3 | 1.52 | 0.85 | 0.013 | 0.0035 | -0.0095 | -73 |
| | | | | | | | **Mean %$\epsilon$:** | **-50** |
| | | | | | | | **RMSE:** | **0.044** |

*Table 10.3b: A selection of test results (MF soils) for an ANN ensemble where $K_{FC}$ and $\hat{K}_{FC}$ represent expected and ensemble-predicted values of the soil hydraulic conductivity, respectively, and $\hat{K}_{FC} - K_{FC}$ represents the prediction error. The percentage error (%$\epsilon$), mean percentage error and the RMSE of the ten predictions are also shown*

## 10.4 Discussion

It is generally acknowledged in the research community that hydraulic conductivity is significantly more difficult to model than water retention using PTFs, regardless of the methodology used or the value of pressure head investigated. One possible reason for this is that, considering two soil horizons having identical soil textural composition, the two horizons may well have very different values of hydraulic conductivity (at a given pressure head), due to the fact that the textural composition confers no information about the size, extent or distribution of the pores. In dry weather, clayey soil may crack, allowing direct hydraulic routes through the soil, whereas the same dry clayey soil that has not cracked will not necessarily have direct hydraulic channels, and will therefore conduct comparatively less water. For water retention, the soil textural composition tells the modeller how much pore space there is in a given soil horizon, and therefore how much water it may hold. For hydraulic conductivity, what is important is not the amount of pore space, but its distribution.

Knowing the soil textural composition will allow a determination of the average pore space, and therefore the average hydraulic conductivity for a given soil horizon. With sufficient data, the modeller may therefore build a PTF model of the average hydraulic conductivity, thus, despite soil textural composition conferring no information about the distribution of the pore space, it is reasonable to postulate that hydraulic conductivity may be successfully modelled by PTF methods.

Many different approaches have been used here in an attempt to successfully construct an ANN-PTF ensemble of hydraulic conductivity. Such approaches include increasing the number of soil horizons in each member ANN, the number of ensemble members, and increasing the number of input parameters to six, using predictions from the continuous ANN-PTF of $\theta_{FC}$ as an additional input to the model. None of these methods made the slightest difference to the predictive abilities of the models, both of class and continuous ANN-PTFs (hence, these results are not detailed further). Thus, there are two possible conclusions that one may make from these results:

- HYPRES contains insufficient hydraulic conductivity data to be able to effectively construct (to within a specified degree of accuracy) an ANN-PTF model of $K_{FC}$ by the ensemble method.
- It may not be possible to model $K_{FC}$ using SSC, BD, %OM and $\theta_{FC}$ as the input parameters to the model.

The original hypothesis for the research conducted into the modelling of $K_{FC}$ was that it *is* possible to model $K_{FC}$ by the ensemble method. Assuming that this hypothesis is correct, there must exist a minimum amount of data that is required for the model to capture the inherent complexities of the data to within a given degree of accuracy. Finding this minimum data requirement would therefore prove the hypothesis correct. The tests conducted here were constructed around this hypothesis, and the results obtained showed that insufficient data exists to determine this minimum data level, and the hypothesis thus remains unproven.

An alternative, opposite, hypothesis could be raised, stating that it is *not* possible to model $K_{FC}$ by the ensemble method (perhaps because of the inherent physical properties of soil). For this hypothesis to be correct, one must define a maximum amount of data. Below this level the model fails to capture the complexities of the data, whilst above this level the amount of data required is too large for the model to be financially feasible (recall that PTFs are inexpensive substitutes for direct measurement of hydraulic properties).

In order to prove the affirmative hypothesis, a minimum amount of data is required, whereas for the negative hypothesis, a maximum amount of data is required. Since I have shown that insufficient data exists to prove that it *is* possible to successfully model $K_{FC}$ by the ensemble method, it follows, therefore, that the amount of HYPRES data available is also insufficient to prove that it is *not* possible to successfully model $K_{FC}$ by the ensemble method.

The difficulties that I have encountered in constructing ensemble PTFs of hydraulic conductivity concur with the findings of Batjes (1996), who stated that '*The number of measured K(h) data in WISE is still too limited for developing pedotransfer*

*functions for K(h). More suitable databases for computing such relationships are being developed elsewhere'*. WISE has 4,353 soil horizons.

Since no evidence or conjecture to the contrary exists, either in this thesis or in current literature, I will, for the time being, continue with the hypothesis that it is possible for $K_{FC}$ to be modelled by ANN-PTF ensemble methods. My conclusion is thus that the current version of HYPRES contains insufficient data to allow modelling of hydraulic conductivity.

The USDA-NRCS database does not contain measurements of hydraulic conductivity.

## Chapter Conclusions

Modelling of hydraulic conductivity, by whatever means, is acknowledged by the research community to be more difficult than modelling of soil water retention. This is due to the fact that it is the size, extent and distribution of the pore spaces that is important in any determination of hydraulic conductivity, and the soil textural composition confers absolutely no information about these. Despite this, the soil textural composition tells us the *average* pore space. Thus, the *average* hydraulic conductivity (at a given value of pressure head) may be modelled, assuming that sufficient data exists.

Constructing an MF class ANN-PTF ensemble of hydraulic conductivity at field capacity, using ten ensemble members and 70 MF soil horizons per ANN member, resulted in a model that produces predictions that are typically one order of magnitude different from the expectation (mostly underestimating the hydraulic conductivity). Increasing the number of soil horizons per ANN member and the number of ensemble members did not result in an improvement to the model's predictive abilities. Additionally, using predictions of water retention at field capacity as a supplementary input to the ensemble model also had no effect on its predictive abilities.

Since there is no evidence, anecdotal or otherwise, that hydraulic conductivity cannot be modelled by PTFs, I have concluded that the current version of HYPRES contains insufficient data to allow successful modelling of hydraulic conductivity at field capacity by the ANN-PTF ensemble method described here.

# Chapter 11

# Discussion and Conclusions

For all the artificial neural network pedotransfer function research that has been reported in the literature, seemingly little, if any, time has been spent in determining how much data is required for stringent training and testing. The rationale behind the research reported in this thesis is that, in order to construct such pedotransfer function models one needs to consider the needs of the model and the data. There exists in any modelling methodology a trade-off between the *amount* of data used and the *value* of that data, and this point is addressed in this thesis by attempting to construct models that are optimised for the data such that the maximum amount of value is extracted from the minimum model data requirements. In order to achieve this, all parts of the modelling procedure have been examined in detail, test models have been constructed and tested, and conjectured improvements implemented in an iterative process. Optimised models have been constructed from these conjectured improvements, and have been tested in detail. The outcomes of all of these suggested improvements and of the final models are discussed in detail in this chapter. Conclusions of the research conducted are also detailed in this chapter.

## 11.1 Pedotransfer Functions

The availability of fresh, clean water is critical to the survival of the human species. Most of the water that he ingests is locked in agricultural products (crops, such as vegetables and grain) and livestock (meat and dairy products), and these rely on water that is stored within the soil and is available for extraction by plant life. The quality of the soil as a store of root-accessible water depends essentially on the size of the pore spaces, and the pore space is the medium that stores and conducts water. It follows then, that the interactions between water and soil (the soil hydraulic properties) should be characterised in some detail in order for the flux (hydraulic conductivity) and storage (water retention) of water at the land surface to be sustainably managed and utilised. The soil hydraulic properties, therefore, play a crucial role in models dealing with the retention and transport of water and solutes.

However, soil hydraulic properties are very time-consuming and expensive to measure by direct means. Conversely, routinely collected soil survey parameters, such as soil texture, dry bulk density and organic matter content, are relatively cheap and

easy to collect. The solution to this disparity is to transfer the data we *have* (soil survey parameters) into the data we *need* (soil hydraulic properties), and mathematical regression models called pedotransfer functions have been developed to achieve this. Pedotransfer functions are the means by which soil scientists may indirectly estimate the latter from the former, and they are often calculated with the aid of one of the many databases of soil hydraulic properties available to soil scientists. Different types of pedotransfer functions, artificial neural network methods in particular, have become important tools in research directed towards quantifying the most important physical and biological processes active in saturated and unsaturated soils, providing a measure of correspondence between measured and simulated functional soil behaviour.

## 11.2 Artificial Neural Network Ensembles

In terms of artificial neural network pedotransfer functions it has become relatively easy for modellers to pass most of the available data to the artificial neural network in the training phase, setting aside a portion of the data for testing. This method assumes that the data used for training contains the relevant information to accurately and reliably represent the underlying function of the data. It also assumes that the data used for testing is sufficient to quantitatively determine the precision, resolution and quality of the model. Also, this model construction method violates the principle of Occam's razor, which states that simpler models should be preferred to more complex models. Clearly, these assumptions need to be challenged.

The research detailed in this thesis has been focussed towards determining and challenging the nature of various assumptions in artificial neural network pedotransfer function models, in order to stimulate further research and enhance the quality of future pedotransfer function models. This has involved modelling pedotransfer functions with artificial neural network ensembles. Essentially, ensembles are a collection of individual artificial neural networks that each provide a solution to the same problem, and in this way, the combined strengths of the individual models are augmented, whilst the weaknesses are dimished. Modelling with ensembles involves sub-dividing the available data so that each individual member of the ensemble has some knowledge of all parts of the dataspace. This leads to the obvious question of how best to sub-divide the data to achieve this. In addition, in order to fully

understand the capabilities of the model, the dataset that is used for testing must also have sufficient knowledge of all parts of the dataspace, and be in adequate supply to fully express this. Also, since each ensemble member is required to have some knowledge of all parts of the dataspace, it is pertinent to ask how much knowledge each should have in order to adequately express the complexity of the problem. All of these questions, and more, have been addressed in this thesis.

## 11.3 The HYPRES Database

The HYPRES database has provided the material that I have used in conducting my research. In HYPRES there are 1,791 soil profiles with a total of 5,560 horizons, and these soil horizons are subdivided into 6 FAO (FAO 1990) texture classes (5 mineral and 1 organic). The mineral classes (defined here to have an organic matter content of less than 10%) are termed coarse (C), medium (M), medium fine (MF), fine (F) and very fine (VF). The C soils are composed mainly of large soil particles (mostly sand), with successive soil texture classes being composed of an increasingly greater fraction of smaller particles, where VF soils are composed mainly of small soil particles (mostly clay). In this way, the HYPRES database has natural divisions incorporated into it. Each soil horizon in HYPRES is detailed by information such as soil textural composition (proportion of sand, silt and clay), dry bulk density and organic matter content, along with other routinely collected soil survey parameters. In addition, the measured soil hydraulic properties (water retention and hydraulic conductivities) of each of these soil horizons are detailed by pressure head value. Not all of these parameters have been collected for all soil horizons in HYPRES, and so not all of the 5,560 soil horizons can be used in the modelling process.

Typically, artificial neural network pedotransfer function modellers use some combination of proportions of sand, silt and clay, dry bulk density and organic matter content as input parameters, and water retention (or hydraulic conductivity) at field capacity and/or permanent wilting point as input parameter(s). Since data is (relatively) plentiful in HYPRES, I decided that all five of these soil survey parameters should be used as artificial neural network input parameters, and that water retention at field capacity should be used as the output parameter. Selection of soil horizons for the modelling process is conditional in that only mineral soil horizons should be used, and that all soil horizon duplicates should be deleted. On

extracting data from HYPRES, subject to these conditions, 2,779 soil horizons remain available for modelling. Approximately one-third and one-quarter of the selected soil horizons belong to the M and C texture classes, respectively, whilst just over and just under one-fifth belong to the MF and F texture classes, respectively. The remaining 4% belong to the little-represented VF texture class.

## 11.4 The Bias/Variance Trade-Off

In the construction of a continuous PTF, if training data were to be chosen randomly from the 2,779 horizons extracted from HYPRES, we would expect that there would typically be nine times as many M horizons as VF horizons. This would lead to a clear bias in the model towards M soils (as compared with VF soil horizons).

In terms of individual artificial neural networks, there is a minimum level of error that may be attained, and this minimum error is a trade-off between bias and variance. It is a trade-off because attempts to decrease the bias (by adding data samples to the model from under-represented regions of the dataspace) are likely to result in higher variance (more variety in the data), and *vice versa*. An artificial neural network that is set up to generalise well (an optimal compromise between low variance and low bias) must take sufficient account of the data to avoid bias towards a particular sub-set. This is achieved by choosing data in equal proportions from all areas of the dataspace, thus, to exploit the natural divisions in HYPRES, an equal number of soil horizons should be chosen from each of the mineral texture classes to reduce bias towards any particular texture class.

In terms of artificial neural network ensembles, the effect of combining the individual members is usually the result of a reduction (if not complete elimination) in variance, whilst leaving the bias unaltered. Therefore, an effective approach is to take more account of the available data to create a set of artificial neural networks that exhibit high variance, but low bias, since the variance component can be removed by combining. Ensembles can thus provide a way of reducing overall prediction errors by *using* the bias/variance trade-off, rather than being *restricted* by it.

Creating ensemble members is essentially an exercise in increasing variance (and therefore reducing bias). High variance ensemble members can be created by varying the architecture, training algorithm, the training data and/or the initial random weights and biases from which each artificial neural network is trained. Ensemble members in

this thesis have been created by varying the training data (Bootstrapping) and the initial weights and biases.

Combining ensemble members is essentially an exercise in eliminating variance. In general, artificial neural networks that have converged to local minima will perform poorly in different regions of feature space and thus their error terms will not be strongly correlated. It is this lack of correlation that drives the combining method. Thus, combining artificial neural networks efficiently utilises the local minima that other techniques (including single artificial neural network methods) try to avoid. Several methods exist for combining the outputs of artificial neural network members, most of which are beyond the scope of this thesis. Ensemble members in this thesis have been combined by weighted averaging of the individual member outputs.

In general, when creating ensemble members we should attempt to decrease bias as much as possible. This means that, not only should we choose data from all areas of the dataspace, we should use as much of it as possible in the modelling process, since the less bias we have, the more variance we have. When creating ensemble members in real-world problems, the modeller rarely has the luxury of unlimited data, and so has to attempt to determine an optimum balance between achieving low bias and conserving data (a trade-off between the amount of data acquired and the value of that data).

## 11.5 Continuous HYPRES ANN-PTF Ensemble

In this thesis I have constructed a HYPRES continuous artificial neural network ensemble pedotransfer function of water retention at field capacity using soil textural composition, dry bulk density and organic matter content as the input parameters. This section details the model parameters used, reviews the model evaluation criteria, and discusses the modelling results, including results of sensitivity analyses of the model. The two competing error prediction methods detailed in the thesis are also discussed here, along with the results of their application.

### 11.5.1 Modelling Parameters

I determined empirically that there should be ten ensemble members, each with a 5-1-1 architecture and trained with 25 soil horizons Bootstrapped from each of the five texture classes. The inputs were pre-processed and passed to the individual

artificial neural networks, which were then trained to optimisation using the Bayesian Regularisation algorithm. This algorithm has advantages over other algorithms in that it is model-driven, i.e. it is unique in its ability to model not the data, but the *function* that underlies the data. In this sense, it automatically prunes the network so that the architecture matches the ideal model for the data presented, and does not require a validation dataset. When the optimum model (weights, biases and architecture) is found for the data, training is stopped automatically, and the artificial neural network is ready to be challenged.

### 11.5.2 Model Evaluation Criteria

On testing pedotransfer functions, the most common measures of performance are the RMSE (a measure of the mean prediction error) and the ME (a measure of the size and sign of the bias), and these standard measures are adopted in this thesis. It is also standard practice when quoting performance measures for pedotransfer functions to indicate the degree of correlation between the training and test datasets. For individual artificial neural network pedotransfer functions, the available data is typically split so that the training and test datasets are completely independent from each other. In this way, test data that has not been used in the training set is termed as 'unseen', and results of these tests are termed as the 'reliability' of the model. The pedotransfer function may also be tested with the same dataset that was used in the training process, and this data is termed as 'seen', and results of these tests are termed as the 'accuracy' of the model. In terms of individual artificial neural networks, data is either seen or unseen. However, in ensembles there exists a third class of data; that is, data that is seen by one or more members of the ensemble but unseen by the other members. This data I have termed as 'green', referring to the fact that the data has been recycled (used in both the training and test phases of the modelling process), and results of these tests are termed as the 'durability' of the model. I have established that, provided that the number of ensemble members is large and the pool of data to be Bootstrapped from is large, the data that is used in both training and test datasets may still be considered to be independent, and data may be conserved by recycling it. I have also established that, for the data modelled here, the reliability, durability and overall results of a given ensemble pedotransfer function are virtually identical, and hence, results of testing may be described in terms of reliability, durability or overall, with no loss of precision. All test results are thus quoted as the overall results.

Additionally I have determined empirically that there should be at least 150 test horizons available per texture class to ensure that tests are statistically stringent and reliable to two decimal points.

### 11.5.3 Discussion of Results

In total ten HYPRES continuous artificial neural network ensemble pedotransfer functions of water retention at field capacity were constructed, and their RMSEs showed a remarkable stability, varying little in terms of their predictive abilities, overall and for each texture class. In addition, all MEs were at least one order of magnitude smaller than the corresponding RMSE results, indicating that the bias in each of the models is very small. It was also encouraging to note that the sign of the ME is not consistent, in other words, field capacity water retention is neither consistently under- nor over-estimated. These ensemble results were compared with results that were obtained by constructing a single artificial neural network pedotransfer function. The comparisons show that an ensemble requires approximately one-third of the amount of soil horizons to provide equivalent results to the single artificial neural network method. Since pedotransfer functions constructed from local data perform better than those built from data spread from a wider area, this is a crucial result. Typically, local data is less bountiful, and hence data conservation becomes a critical factor in artificial neural network pedotransfer functions. Ensembles appear to be the answer to the problem since they allow data to be recycled.

### 11.5.4 Sensitivity Analyses

Sensitivity analyses of the ensemble model can help inform field scientists which particular parameters are of most importance and should be measured with most care and diligence. Results of such analyses of the HYPRES continuous artificial neural network ensemble pedotransfer function of water retention at field capacity show that, for single parameter omissions, only dry bulk density has a significant effect, worsening the predictive ability of the ensemble. On omitting dry bulk density and organic matter content, the predictions deteriorate, and the deterioration increases with the omission of soil textural composition (only dry bulk density and organic matter content used in the model). This is the case with all texture classes.

### 11.5.5 Error Prediction Methods

One of the key requirements of any statistical analysis system should be to assess its own confidence in a decision and estimate probable error bars. In this thesis I have presented two possible ways of producing errors of individual predictions – a Bootstrapping method, and an error ensemble method.

#### Bootstrapping Method

When a single artificial neural network is used to make predictions, confidence may not be determined for individual results. However, when a number of training sets are Bootstrapped from the available data and trained with individual artificial neural networks, the spread of predictions (calculated as the standard deviation) will give a measure of confidence in those predictions. This is termed the Bootstrapping method of error prediction, and provides a measure of the extent of agreement between the individual artificial neural networks in the ensemble.

#### Error Ensemble Method

The error ensemble method of error prediction involves taking the predictions of the ensemble pedotransfer function and comparing them to the expected values. The variance of the predictions can then be calculated, and can be used (in a pre-processed form) as the output parameter to an error ensemble that uses precisely the same input parameters and construction techniques as the ensemble pedotransfer function. To recap, the ensemble pedotransfer function uses proportions of sand, silt and clay, dry bulk density and organic matter content as the input parameters, and their respective measurements of water retention at field capacity as the output parameter. The error ensemble uses these same five soil survey parameters as ensemble input parameters, and their respective errors (in the form of the pre-processed variance) as the output parameter. The error ensemble thus provides a measure of the discrepancy between the ensemble prediction and the expectation.

#### Discussion of Results

On determining the statistical properties of the various parameters used, I determined that, in general, as the dry bulk density of soil samples increases, the water retention at field capacity decreases, and also that as the organic matter content of soil samples increases, the water retention at field capacity increases. In addition,

although the range of values for dry bulk density in HYPRES is [0.54, 1.97], the majority of M textured soil horizons fall in the range [1.4, 1.8], with a mean value of 1.49. Thus, in the range [0, 2] it is expected that M soils that have a high value for dry bulk density should be predicted with the highest accuracy, and thus highest confidence (lowest errors). In terms of the organic matter content, the range of values in HYPRES is [0, 9.9]. However, the majority of values fall in the range [0, 3], with a mean value of 1.6 (median of 1.22) for M textured soil horizons. Thus, in the range [0, 10] it is expected that M soils that have a low value for organic matter content should be predicted with the highest accuracy, and thus highest confidence (lowest errors).

Tests of a given M soil horizon show that for the HYPRES continuous artificial neural network pedotransfer function ensemble of water retention at field capacity, the Bootstrapping and error ensemble methods produce similarly confident results in the region of high dry bulk density and low organic matter content, as expected. Outwith these regions, confidence in predictions varied significantly, and this is also expected, since data in these outlying regions is comparatively scarce, and the ensemble has less knowledge of the dataspace, and is thus being asked to extrapolate – a task that artificial neural networks are notoriously poor at performing.

## 11.6 Class HYPRES ANN-PTF Ensemble

In this thesis I have also constructed a HYPRES class artificial neural network ensemble pedotransfer function of water retention at field capacity for M textured soil horizons using precisely the same methods as were used in the continuous model. The only exception to this is that I determined empirically that there should be ten ensemble members, each with a 5-1-1 architecture and trained with 100 M soil horizons Bootstrapped from the available data. This section details the results of tests conducted on this model, including results of sensitivity analyses and error prediction methods.

### 11.6.1 Discussion of Results

In total ten HYPRES class artificial neural network ensemble pedotransfer functions of water retention at field capacity for M soils were constructed, and their RMSEs showed a remarkable stability, varying little in terms of their predictive

abilities for the M texture class. In addition, all MEs were at least one order of magnitude smaller than the corresponding RMSE results, indicating that the bias in each of the models is very small. It was also encouraging to note that the sign of the ME is not consistent, in other words, field capacity water retention is neither consistently under- nor over-estimated.

### 11.6.2 Sensitivity Analyses

Results of sensitivity analyses of the HYPRES class artificial neural network ensemble pedotransfer function of water retention at field capacity for M soils show that omitting an individual parameter has no significant effect on the model as a whole. On omitting dry bulk density and organic matter content, the predictions deteriorate, and the deterioration increases with the omission of soil textural composition (only dry bulk density and organic matter content used in the model).

### 11.6.3 Error Prediction Methods

Tests of a given M soil horizon show that for the HYPRES class artificial neural network pedotransfer function ensemble of water retention at field capacity for M soils, the Bootstrapping and error ensemble methods produce similarly confident results in the region of high dry bulk density and low organic matter content, as expected. Outwith these regions, confidence in predictions is lower, particularly with respect to the organic matter content, and this is also expected, since data in these outlying regions is comparatively scarce, and the ensemble has less knowledge of the dataspace, and is thus being asked to extrapolate.

## 11.7 Continuous USDA-NRCS ANN-PTF Ensemble

To recap, I have constructed both continuous and class (M soils only) artificial neural network ensemble pedotransfer functions of water retention at field capacity using data from the HYPRES database. In addition I have constructed two methods of error prediction, the Bootstrapping and the error ensemble methods, and these have both been applied to the class and continuous ensemble pedotransfer functions. To assess whether the methods, tests and results that were developed using HYPRES are repeatable, I have applied all of these models to the USDA-NRCS database. Results of tests conducted are discussed here, along with results of sensitivity analyses and error prediction.

### 11.7.1 Discussion of Results

In total ten USDA-NRCS continuous artificial neural network ensemble pedotransfer functions of water retention at field capacity were constructed, and their RMSEs varied little in terms of their predictive abilities, overall and for each texture class. All MEs were at least one order of magnitude smaller than the corresponding RMSE results, indicating that the bias in each of the models is very small. In addition, the sign of the ME is not consistent, indicating that field capacity water retention is neither consistently under- nor over-estimated.

### 11.7.2 Sensitivity Analyses

Results of sensitivity analyses of the USDA-NRCS continuous artificial neural network ensemble pedotransfer function of water retention at field capacity show that omission of any single parameter from the model has no significant effect on predictive ability of the ensemble. Omitting dry bulk density and organic matter content results in a small decrease in predictive ability for the M and VF texture classes, and results in a significant decrease in predictive ability for C soil horizons. Omitting soil textural composition (only dry bulk density and organic matter content used in the model) results in an almost complete failure of the model to predict water retention at field capacity.

### 11.7.3 Error Prediction Methods

On determining the statistical properties of the USDA-NRCS data, I determined that, in general, as the dry bulk density of soil samples increases, the water retention at field capacity decreases, and also that as the organic matter content of soil samples increases, the water retention at field capacity increases. In addition, although the range of values for dry bulk density in USDA-NRCS is [0.13, 2.16], the majority of M textured soil horizons fall in the range [1.1, 1.7], with a mean value of 1.31 (median of 1.35). Thus, in the range [0, 2] it is expected that M soils that have a high value for dry bulk density should be predicted with the highest accuracy, and thus highest confidence (lowest errors). In terms of the organic matter content, the range of values in USDA-NRCS is [0.01, 9.97]. However, the majority of values fall in the range [0, 3], with a mean value of 1.88 (median of 1.35) for M textured soil horizons. Thus, in the range [0, 10] it is expected that M soils that have a low value for organic

matter content should be predicted with the highest accuracy, and thus highest confidence (lowest errors).

Tests of a given M soil horizon show that for the USDA-NRCS continuous artificial neural network pedotransfer function ensemble of water retention at field capacity, the Bootstrapping and error ensemble methods produce similarly confident results in the region of high dry bulk density and low organic matter content, as expected. Outwith these regions, confidence in predictions is lower, particularly with respect to the organic matter content, and this is also expected, since data in these outlying regions is comparatively scarce, and the ensemble has less knowledge of the dataspace, and is thus being asked to extrapolate.

## 11.8 Class USDA-NRCS ANN-PTF Ensemble

I have also constructed a USDA-NRCS class artificial neural network ensemble pedotransfer function of water retention at field capacity for M soils using soil textural composition, dry bulk density and organic matter content as the input parameters. The model has been implemented in precisely the same way as was the HYPRES class pedotransfer function model. Results of tests conducted on this model are discussed here, including results of sensitivity analyses and error prediction.

### 11.8.1 Discussion of Results

In total ten USDA-NRCS class artificial neural network ensemble pedotransfer functions of water retention at field capacity were constructed, and their RMSEs varied little in terms of their abilities to predict water retention at field capacity for M soil horizons. In addition, all MEs were at least one order of magnitude smaller than the corresponding RMSE results, indicating that the bias in each of the models is very small. Moreover, the sign of the ME is not consistent, in other words, water retention at field capacity is neither consistently under- nor over-estimated.

### 11.8.2 Sensitivity Analyses

Results of sensitivity analyses of the USDA-NRCS class artificial neural network ensemble pedotransfer function of water retention at field capacity for M soils show that omission of any single parameter from the model has no significant effect on predictive ability of the ensemble. In terms of multiple parameter omissions, omitting dry bulk density and organic matter content has no significant effect. However,

omitting soil textural composition (only dry bulk density and organic matter content used in the model) results in a significant deterioration in predictive ability of the ensemble.

### 11.8.3 Error Prediction Methods

Tests of a given M soil horizon show that for the USDA-NRCS class artificial neural network pedotransfer function ensemble of water retention at field capacity for M soils, the Bootstrapping and error ensemble methods produce similarly confident results in the region of high dry bulk density and low organic matter content, as expected. Outwith these regions, confidence in predictions is lower, particularly with respect to the organic matter content, and this is also expected, since data in these outlying regions is comparatively scarce, and the ensemble has less knowledge of the dataspace, and is thus being asked to extrapolate.

## 11.9 Model Comparisons

Since the model used to construct a continuous artificial neural network pedotransfer function ensemble of water retention at field capacity remains unchanged for both the HYPRES and USDA-NRCS datasets, the results may be compared to assess the suitability of each for such a model. In this section, a discussion of direct comparisons between class and continuous models of HYPRES and USDA-NRCS datasets are made, including error prediction methods applied to such.

### 11.9.1 HYPRES Class and Continuous PTFs

In order to assess the differences in the class and continuous ensemble pedotransfer function models I have directly compared the M texture class sensitivity analyses results of such, conducted on the HYPRES dataset. The continuous model produces slightly improved predictions of M soil horizons over the class model when all input parameters are used (no parameter omissions), although this improvement is not significant. In terms of single parameter omissions, there are no significant differences in the class and continuous model results. When dry bulk density and organic matter content are omitted, the class model performs slightly better than the continuous model, and significantly better when soil textural composition (only dry bulk density and organic matter content used in the model) are omitted. Perhaps it is reasonable to expect that the class pedotransfer function should produce improved

predictions of M soil horizons over the continuous pedotransfer function, since the class pedotransfer function is a specialised model, using more data in the training phase of the modelling process. Although this is not what the results suggest, it may be inferred that the class model is overall more robust, since it is able to withstand the loss of multiple parameters better than the continuous pedotransfer function.

Contrary to the findings of Wösten *et al.* (1998), who found that their regression-based class pedotransfer functions were significantly more accurate than their continuous pedotransfer functions, the class and continuous ensemble models constructed here produce results of similar accuracy, both to each other and to those found in the current literature.

### 11.9.2 USDA-NRCS Class and Continuous PTFs

Comparing the M texture class sensitivity analyses results of the class and continuous models conducted on the USDA-NRCS dataset, there are no differences in the predictive abilities of the models, regardless of any parameter omissions, with the exception of the omission of soil textural composition (only dry bulk density and organic matter content used in the model). Depriving the USDA-NRCS models of all textural information results in unreliable models. Clearly, models based on the USDA-NRCS dataset are very sensitive to the removal of soil textural information.

Wösten *et al.* (1998) raised the supposition that artificial neural networks may yield more accurate continuous pedotransfer functions than regression-based models. It has been established here that the ensemble method of water retention pedotransfer function construction yields class and continuous models of comparable quality, when HYPRES data are used. Using identical criteria and methods, class and continuous ensemble models using USDA-NRCS data are also of comparable quality. This provides a confirmation that the similarities of the class and continuous model results are not emergent phenomena of the HYPRES dataset, but rather are the consequence of using the ensemble method. It may be concluded, therefore, that the ensemble method yields class and continuous pedotransfer functions of water retention that are of sufficient quality.

### 11.9.3 HYPRES and USDA-NRCS Continuous PTFs

For the continuous model, HYPRES data produced better predictions than USDA-NRCS data for all texture classes, except for F soil horizons, where results

were comparable. Exceptions to this are found when sensitivity analyses are performed, and these show that USDA-NRCS data performs better than HYPRES data in the MF texture class when dry bulk density and organic matter content are omitted, and in the F texture class when dry bulk density, and dry bulk density and organic matter content are omitted. However, the results from both datasets used show that the ensemble model is a very effective method of constructing continuous pedotransfer functions of water retention at field capacity.

### 11.9.4 HYPRES and USDA-NRCS Class PTFs

For the class model for M soil horizons, the class pedotransfer function is more reliable when HYPRES data is used, irrespective of parameter omissions, but particularly so in the case of the omission of all soil textural information. The HYPRES dataset is more sensitive than USDA-NRCS with respect to the omission of dry bulk density, and dry bulk density and organic matter content from the class pedotransfer function ensemble model for M textured soils.

All models, whether class or continuous, or HYPRES or USDA-NRCS showed that as dry bulk density increases water retention at field capacity decreases, and that as organic matter content increases water retention at field capacity increases.

### 11.9.5 HYPRES Class and Continuous Error Models

For the HYPRES error ensemble, both class and continuous models have a flat error response, although both also have some anomalous features, particularly with respect to the continuous model, whilst for the Bootstrapping model, both class and continuous pedotransfer functions produce similar predictions of error

### 11.9.6 USDA-NRCS Class and Continuous Error Models

The USDA-NRCS error ensemble class and continuous models both have smooth error functions, although the continuous model is more sensitive to changes in BD than the class model, whilst for the Bootstrapping model, the continuous pedotransfer function produces error predictions that are twice as large as those predicted by the class pedotransfer function. In addition, the USDA-NRCS Bootstrapping method performed on the class pedotransfer function is independent of BD.

Regardless of which error prediction method is used, all produce similarly confident results in the region of high dry bulk density and low organic matter content, as expected.

## 11.10 Modelling Hydraulic Conductivity

In Chapter 0 of this thesis, I stated that one of the aims of my research was to construct class and continuous artificial neural network ensemble pedotransfer functions of soil hydraulic conductivity at field capacity using data from the HYPRES dataset. Significant amounts of work have been conducted in trying to achieve these specific aims, and although I have been successful in constructing a class model using MF soils, the test results on this model have proven to be very disappointing. On inspection of the results, I found that the RMSEs and MEs of each of the individual artificial neural networks, and of the several ensembles constructed, were almost identical. On closer inspection, I noticed that predictions were typically an order of magnitude smaller than the expected values. Many different approaches were used to try to affect this problem, such as increasing the number of soil horizons, the number of ensemble members, and even increasing the number of input parameters to six, using predictions from the continuous field capacity pedotransfer function of water retention as an additional input to the model. Nothing made the slightest difference to the predictive abilities of the models. Despite being unable to conclude that it possible to model hydraulic conductivity by the ensemble method, I am also unable to conclude that it is not possible. Since no evidence or conjecture to the contrary exists, either in this thesis or in current literature, I will, for the time being, continue with the hypothesis that it is possible to model hydraulic conductivity with the ensemble method. My conclusion, therefore, is that the current version of HYPRES contains insufficient hydraulic conductivity data to be able to be effectively modelled by the ensemble method.

## Chapter Conclusions

There are a number of conclusions that may be drawn from the research reported in this thesis:

- Ensembles are able to produce pedotransfer function models that are more reliable, robust, stable and unbiased than single artificial neural network models.
- Care must be taken to ensure that bias in the dataset is minimised in order to make the best use of the benefits on offer by the ensemble method.
- Continuous artificial neural network pedotransfer function ensembles require one-third or less of the amount of data than do single artificial neural network models to achieve the same results.
- Data may be recycled when using ensemble models, since training and test data may be considered to be independent when the data pool and number of ensemble members are both sufficiently large.
- Ensembles require new terms to describe recycled data (data that is seen by one or more ensemble members, but is unseen by the rest). Data that is both seen and unseen may be termed as 'green' with reference to the fact that it has been recycled, and tests with such data may be referred to as the 'durability' of the ensemble.
- Approximately 150 soil horizons per texture class are required for stringent testing of the water retention ensemble (or indeed any individual water retention artificial neural network).
- 25 soil horizons per texture class are sufficient to allow reliable modelling of continuous pedotransfer functions of water retention at field capacity by the ensemble method.
- Ten ensemble members are sufficient to model pedotransfer functions of water retention at field capacity, regardless of whether it is a class or continuous pedotransfer function that is being modelled.
- The ensemble method produces very effective and stable continuous pedotransfer function models of water retention at field capacity.
- The ensemble method also produces very effective and stable class pedotransfer function models of water retention at field capacity.

- Class and continuous pedotransfer functions have similar accuracies when modelled by the artificial neural network ensemble method.
- 100 soil horizons from the M soil texture class are sufficient to allow reliable modelling of class pedotransfer functions of water retention at field capacity of M soils by the ensemble method.
- The HYPRES and USDA-NRCS databases are good media with which to construct class and continuous PTFs of water retention at field capacity.
- Better class and continuous pedotransfer function ensembles of water retention at field capacity may be built if more soil horizons and ensemble members are used in the training process. However, the bias/variance trade-off remains critical in ensembles, and efforts must still be made to ensure that bias and noise are not introduced into the modelling process by choosing data carelessly.
- It should be noted in the ensemble methods constructed herein that since data has been chosen carefully, bias in the models is small, and often close to zero. Due to the bias/variance trade-off, when combining ensemble members variance is significantly reduced, if not completely eliminated. Therefore, the greatest contribution to the errors in the ensemble models comes from noise. Noise in the data invariably originates from the effects of hysteresis, errors in measurement, fitting errors, calculation or insertion into the database and disturbance of soil samples prior to determination of water retention measurements. Thus, databases of soil hydraulic parameters are only as valuable as the quality of data that resides within them. Clearly, great care should be taken to ensure that *all* data in databases is as noise-free as possible to ensure that reliable modelling can be performed.
- Sensitivity analyses allow modellers to inform field scientists which data should be measured with greatest care. Armed with this knowledge, field scientists can then gather better quality data, ensuring that the data in the databases is of high quality and possesses as little noise as possible. This high quality data may then be used by the modeller to construct high quality models, which are then able to more accurately predict the sources of error in the data. Thus, modellers and field scientists may work more closely together to produce synergistic effects in the modelling process.

- Both Bootstrapping and error ensemble methods are valid methods of error prediction, albeit that they predict slightly differing types of error. The Bootstrapping method measures the extent to which the individaul ensemble members agree on the prediction. The error ensemble method measures the discrepancy between the ensemble prediction and the expectation.
- Predictions should only be made from ensembles if they are asked to interpolate. Extrapolation is not recommended, and confidence should be very low in regions of the dataspace where extrapolation is required, whichever method of error prediction is used.
- Insufficient data exists in the current version of HYPRES with which to successfully construct an artificial neural network ensemble pedotransfer function of hydraulic conductivity at field capacity.
- Large databases are not required in order to model reliable class and continuous artificial neural network ensemble pedotransfer functions of water retention at field capacity.
- However, large databases are required in order to fuel good research – this thesis would not exist without the availability of databases of soil hydraulic properties.

If there was a single take-home message from the research conducted in this thesis, for both ANN modellers and soil scientists, it is that ANN ensembles are able to produce both class and continuous PTFs of water retention that are more reliable, robust, stable and unbiased than single ANN-PTFs, particularly when data is scarce.

Finally, it is hoped that the research detailed in this thesis will inspire other environmental modellers to experiment with the ensemble methods. Also, it is hoped that my work will inspire soil scientists to continue work with large databases of soil hydraulic properties, such as HYPRES and USDA-NRCS, and keep them in the public domain so that further research may be carried out by all.

# Chapter 12

# Future Work

In this thesis, focus has been given to the production of pedotransfer functions of water retention at field capacity. Although this is important for modellers of soil hydraulic properties, it is more useful to have a model of water retention at permanent wilting point *in addition* to one at field capacity, so that the water that is available in the soil for extraction by plants may be calculated. Following on from the focus in this thesis, a modeller may construct an ensemble of water retention at permanent wilting point, so that there exists independent ensembles for each of the two water retention curve inflexion points. Additionally, he may model an ensemble that has two outputs – the field capacity and permanent wilting points. In this way, the two different types of ensemble may be compared to determine whether it is more accurate to construct independent ensembles (modules) or a combined ensemble of the inflexion points of the water retention curve.

Also, in this thesis a class pedotransfer function of M soil horizons only has been built. In order to fully compare the effectivenesses of the class and continuous methods, class pedotransfer functions of C, MF, F and VF should be constructed. Each of these ensembles would be a self-contained module, and input test data would require 'switching'. This involves determining which of the class pedotransfer functions is a specialist for the particular test soil horizon. This would probably involve one of two methods – sequential or rule-based switching. In sequential switching a classification artificial neural network would be used in stage 1 to decide to which of the soil texture classes the test soil horizon belongs. The test horizon would then be passed on to the appropriate class ensemble pedotransfer function at stage 2. In rule-based switching, explicit rules are used with which to decide to which soil texture classes the test soil horizon belongs, such as those detailed in Table 5.2d.

Success has been achieved in this thesis in constructing class and continuous artificial neural network ensemble pedotransfer functions of water retention at field capacity. In addition, two different types of error prediction models have been

constructed. However, these have been implemented for point pedotransfer functions only. Sufficient data exists in HYPRES for the same models to be constructed using the function pedotransfer function method. If these were to be created, then point and function pedotransfer functions of the same data and ensemble construction techniques may be directly compared to determine the relative effectivenesses of the point and function pedotransfer function techniques. To achieve this, a generic artificial neural network ensemble model for water retention could be built, perhaps with a user-friendly interface to enable the investigator to concentrate on the pedotransfer function, rather than on the programming aspect of the modelling. With such a generic tool, many different types of investigation could be achieved simply and quickly without the need for re-writing large sections of the pedotransfer function code.

Such investigations could perhaps include independently scaling the input parameters. Recall that each of the five input parameters was re-scaled by Equation 6.1b. The re-scaling algorithm was designed to re-scale proportions of sand, silt and clay from [0, 100] into the range [0, 1]. However, values of dry bulk density naturally fall in the range [0, 2], and so their re-scaled values fall into a much tighter range, and perhaps may not be resolved quite so readily. Re-scaling each of the input parameters individually allows them to all fall naturally into the range [0, 1], and would all be in an optimum condition for input to the artificial neural networks. It may not effect a more accurate model (recall that the error surface is fixed for a given dataset), however, it may allow the artificial neural networks to reach optimisation more quickly.

The FAO (1990) definitions for sand, silt and clay are that clay particles are defined to have a size of less than 2μm, silt particles have sizes in the range 2-50μm and sand particles have a range of 50-2000μm. However, these sizes are chosen arbitrarily, and convey no *functional* information about the soil. It would be interesting to see if particle size analyses could be conducted on the raw data contained in HYPRES by various methods, such as classification artificial neural networks and Bayesian Belief Nets to see if different particle size ranges that convey functional information could be found. These functional ranges would provide a stronger basis for ensembles, since data could be extracted from functional areas of the dataspace to reduce bias. Additionally, alternative methods for data selection

could be investigated, such as Group Methods of Data Handling, regression trees, the different types of soil textural triangle that exist, etc..

Automatic data selection is also an area of research that could be investigated. Consider 25 soil horizons Bootstrapped from a pool of 100 data points. Do these 25 soil horizons best represent the function underlying the data? They may convey a *good* understanding of the function underlying the data, but probably not the *best* understanding, since they were sampled randomly. Automatic data selection would probably involve training an artificial neural network to optimisation with the randomly chosen 25 soil horizons. At this point, a random training horizon would be removed from the training set and replaced by a random horizon from the data pool. The network would then continue the training process from this point (utilising the optimised weights and biases) until a new optimisation is found. If the sum-squared error of the new optimisation is higher than the old optimisation, then the old data point is of higher quality than the new one, and the new data point is returned to the data pool, and replaced in the training set by the old data point. However, if the sum-squared error of the new optimisation is lower than the old optimisation, then the new data point is of higher quality than the old point, and the old point is returned to the data pool. This process would be repeated until the lowest sum-squared error of the data pool is found. In this way, the *best* data to represent the function underlying the data may be found for the artificial neural network, and this process may then be extended to the ensemble model. Alternatively, the use of Bayesian statistics, Genetic Algorithms and Bayesian Belief Nets could be investigated to determine the most appropriate data and size of the training dataset.

In this thesis, a strong emphasis has been given to the advantages gained by using the Bayesian Regularisation training algorithm. This is mostly because it is model-driven, whilst all other algorithms are data-driven. Bayesian statistics may be used to optimise a single artificial neural network model, but can they also be used to optimise an ensemble of artificial neural networks? In this way, the optimum number of members and also the optimum method for combining members could be found automatically.

There have been many different types of pedotransfer function constructed using widely varying methods, such as statistical regression-based methods, regression trees and artificial neural networks. Equation 4.2a in this thesis shows that ensembles of estimators yield significantly better estimates than single estimator methods. This

equation is a very powerful one, since, although in this thesis it has been applied to artificial neural networks, it can actually be applied to *any* type of estimator. Indeed it may be applied to *combinations* of estimators, for instance, the results of artificial neural networks and multiple linear regression may be combined to produce a cross-modal ensemble. Taking the mean of the individual predictions from each of the methods would yield more accurate results than any of the individual methods. In particular, taking the weighted average would yield significantly improved estimates over taking the mean. Of course, this could only be achieved if the mean squared error of the methods used are given. In addition to this, there have been many different types of artificial neural network methods used to construct pedotransfer functions, and each have their own strengths and weaknesses. An ensemble of artificial neural networks using these differing methods may be constructed to create an all-encompassing world pedotransfer function knowledge bank.

When challenging the ensemble and the error prediction methods detailed in this thesis, only a single M soil horizon has been used. To determine degrees of confidence in different regions of the dataspace, the ensembles should be tested with different soil horizons. For instance, when dry bulk density is high, predictions should be accurate and confidence high. Is this the case for all soil horizons in this region of the dataspace? In regions where confidence should be low, is confidence low for all soil samples in this region, or are there exceptions? Is there anything that can be learnt from these exceptions, and if so, can we improve the model from this new information? Only stringent testing of all the models can answer these questions.

In addition, there is another possible method of producing error predictions on individual tests. This is termed the 'Hessian' method of error prediction, and the theory behind this involves performing a transformation on the Hessian matrix (this is used as part of the Levenberg-Marquardt and Bayesian Regularisation algorithmic calculations). An explanation on how this is achieved is beyond the scope of this thesis (the details may be found in MacKay (1992b)), however, this provides another method of error prediction, with which to compare the two methods detailed in this thesis.

# References

Acutis, M. and Donatelli, M., 2003. SOILPAR 2.00: software to estimate soil hydraulic parameters and functions. *European Journal of Agronomy.* 18:373-377.

Ahuja, L.R., Naney, J.W. and Williams, R.D. 1985. Estimating soil water characteristics from simpler properties or limited data. *Soil Science Society of America Journal.* 49:1100-1105.

Ahuja, L.R., *et al.* 1989. Relationship between green and ampt parameters based on scaling concepts and field-measured hydraulic data. *Water Resources Research.* 25:1766-1770.

Aina, P.O. and Periaswamy, S.P. 1985. Estimating available water-holding capacity of western Nigerian soils from soil texture and bulk density, using core and sieved samples. *Soil Science.* 140:55-58.

Alexander, L. and Skaggs, R.W. 1987. Predicting unsaturated hydraulic conductivity from soil texture. *Journal of Irrigation and Drainage Engineering.* 113:184-197.

Al-Ghoneim, K. and Vijaya Kumar, B.V.K. 1995. Learning ranks with neural networks. In: Applications and science of artificial neural networks, Proceedings of the SPIE. 2492:446-464.

Anand, R. *et al.* 1995. Efficient classification for multiclass problems using modular neural networks. *IEEE Transactions on Neural Networks.* 6:117-124.

Anderson, M.G., *et al.* 1985. On soil water retention curves and hydrological forecasting in ungauged catchments. *Nordic Hydrology.* 16:11-32.

Arya, L.M. and Dierolf, T.S. 1992. Predicting soil moisture characteristics from particle-size distributions, an improved method to calculate pore radii from particle radii. In: Van Genuchten, M. T., Leij, F. J. and Lund, L. J. eds. Indirect methods for estimating the hydraulic properties of unsaturated soils. Proceedings of the international workshop on indirect methods for estimating the hydraulic properties of unsaturated soils, Riverside, California, 11-13 October 1989. 115-124.

Arya, L.M. and Paris, J.F. 1981. A physicoempirical model to predict the soil moisture characteristic from particle-size distribution and bulk density data. *Soil Science Society of America Journal.* 45:1023-1030.

Bachmann, J. and Hartge, K.H. 1991. Estimating soil water characteristics obtained by basic soil data-a comparison of indirect methods. *Z. Pflansenernähr. Bodenkd.* 155:109-114.

Bates, J.M. and Granger, C.W.J. 1969. The combination of forecasts. *Operations Research Quarterly*. 20:451-468.

Batjes, N.H. 1996. Development of a world data set of soil water retention properties using pedotransfer rules. *Geoderma*. 71(1-2):31-52.

Beke, G.L. and McCormic, M.J. 1985. Predicting volumetric water retention for subsoil materials from Colchester County, Nova Scotia. *Canadian Journal of Soil Science*. 65:233-236.

Bell, M.A. 1993. Organic matter, soil properties and wheat production in the high Valley of Mexico. *Soil Science*. 156:86-93.

Bell, M.A. and van Keulen, H. 1996. Effect of soil disturbance on pedotransfer function development for field capacity. *Soil Technology*. 8:321-329.

Bernoulli, D. 1738. Opus academicum ab auctore, dum petropoli ageret, congestum, argentorati: Sumptibus johannis reinholdi dulseckeri, typis joh. Henr. Deckeri, typographi basiliensis. *Hydrodynamica, sive De Viribus et Motibus Fluidorum Commentarii.*

Bhatnagar, D., Nagarajarao, Y. and Gupta, R.P. 1979. Influence of water content and soil properties on unsaturated hydraulic conductivity of some red and black soils. *Z. Pflansenernähr. Bodenkd.* 142:99-108.

Bishop, C.M. 1995. Neural networks for pattern recognition. Oxford University Press, Oxford.

Bloemen, G.W. 1980. Calculation of hydraulic conductivities from texture and organic matter content. *Z. Pflansenernähr. Bodenkd.* 143:581-605.

Bouma, J. 1989. Using soil survey data for quantitative land evaluation. *Advances in Soil Science*. 9:177-213.

Bouma, J. and van Lanen, H.A.J. 1987. Transfer functions and threshold values: From soil characteristics to land qualities. In: Beek, K. J., *et al.* Ed. Proc. ISSS/SSSA workshops on quantified land evaluation procedures, Washington, DC, 27 April–2 May 1999. 106-111.

Brooks, R.H. and Corey, A.T. 1964. Hydraulic properties of porous media. *Hydrology*. 3.

Bruand, A. *et al.* 1996. Variabilité des propriétés de rétention en eau des sols, importance de la densité apparante. *Etude et Gestion des Sols*. 3:27-40.

Brutsaert, W. 1967. Some methods of calculating unsaturated permeability. *Transactions of the American Society of Agricultural Engineers*. 10:400-404.

Burdine, N.T. 1953. Relative permeability calculations from pore-size distribution data. *Petroleum Transactions of the American Institute of Mining Engineering*. 198:71-77.

Calhoun, F.G., Hammond, L.C. and Caldwell, R.E. 1972. Influence of particle size and organic matter on water retention in selected Florida soils. *Soil and Crop Science of Florida Proceedings*. 32:111-113.

Carman, P.C. 1937. Fluid flow through a granular bed. *Transactions of the Institution of Chemical Engineers, London*. 15:150-167.

Chester, M. 1993. Neural networks: a tutorial. PTR Prentice-Hall Inc., Englewood Cliffs, New Jersey 07632, 50-53.

Commission of the European Communities (CEC) 1985. Explanatory text and map sheets of the 1: 1.000.000 soil map of the European Communities.

Cornelis, W.M. *et al.*, 2001. Evaluation of pedotransfer functions for predicting the soil moisture retention curve. *Soil Science Society of America Journal*. 65:638-648.

Cunningham, P., Carney, J. and Jacob, S. 2000. Stability problems with artificial neural networks and the ensemble solution. *Artificial Intelligence in Medicine*. 20:217-225.

Danaher, S., Datta, S., Waddle, I. and Hackney, P. 2004. Erosion modelling using Bayesian regulated artificial neural networks. *Wear*. 256:879-888.

Darcy, H. 1856. Les fontaines publiques de la ville de Dijon. Dalmont, Paris.

Demuth, H. and Beale, M. 2001. Neural network toolbox user's guide. The Mathworks, inc., 3 Apple Hill Drive, Natick, MA, 01760-2098.

DTI's Neural Computing Technology Transfer Programme 1994. Best practice guidelines for developing neural computing applications. Crown publications, London.

EAD-ANL. [Online]. Environmental Assessment Division of Argonne National Laboratory. Available from the World Wide Web: URL: http://web.ead.anl.gov/resrad/datacoll/volcont.htm.

Efron, B. 1983. Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*. 78:316-331.

Elsenbeer, H. 2001. Pedotransfer functions in hydrology. *Journal of Hydrology*. 251(3-4):121-122

Fausett, L.V. 1994. Fundamentals of neural networks: architectures, algorithms, and applications. Prentice-Hall Inc., Upper Saddle River, New Jersey 07458, 23-25.

Food and Agriculture Organisation (FAO) 1990. Guidelines for soil description (3rd ed.), FAO/ISRIC, Rome.

Fischer, W. *et al.* 1987. Amelioration of cholinergic neuron atrophy and spatial memory impairment in aged rats by nerve growth factor. *Nature*. 329:65-68.

Franzmeier, D.P. 1991. Estimation of hydraulic conductivity from effective porosity data for some Indiana soils. *Soil Science Society of America Journal*. 55:1801-1803.

Geman, S., Bienenstock, E. and Doursat, R. 1992. Neural networks and the bias/variance dilemma. *Neural Computation.* 4:1-58.

Goutte, C. 1997. Note on free lunches and cross-validation. *Neural Computation.* 9(6):1246-1249.

Green, R.E. and Corey, J.C. 1971. Calculation of hydraulic conductivity, a further evaluation of some predictive methods. *Soil Science Society of America Journal.* Proceedings 35:3-7.

Gregson, K., Hector, D.J. and McGowan, M. 1987. A one-parameter model for the soil water characteristic. *Journal of Soil Science.* 38:483-486.

Grofman, B. and Owen, G. 1986. Editors, Information Pooling and Group Decision Making: Proceedings of the Second University of California, Irvine Conference on Political Economy. JAI Press, Inc., Greenwich, Connecticut.

Gupta, S.C. and Larson, W.E. 1979. Estimating soil water characteristic from particle size distribution, organic matter percent, and bulk density. *Water Resources Research.* 15:1633-1635.

Hagan, M.T., Demuth, H.B., and Beale, M.H. 1996. Neural network design. PWS Publishing Company, Boston, MA 02116.

Haines, W.B. 1930. Studies in the physical properties of soils. V. The hysteresis effect in capillary properties and the modes of moisture distribution associated therewith. *Journal of Agronomic Science.* 20:97-116.

Hansen, L. and Salamon, P. 1990. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 12(10):993-1001.

Hrycej, T. 1992. Modular learning in neural networks - a modularized approach to neural network classification. In Branko Soucek (Ed.): Sixth-generation computer technology series. Wiley-Interscience, 34-35.

Jacobs, R.A. 1995. Methods for combining experts' probability assessments. *Neural Computation.* 7:867-888.

Koekkoek, E. and Booltink, H. 1999. Neural network models to predict soil water retention. *European Journal of Soil Science.* 50:489-495.

Kolen, J.F. and Pollack, J.B. 1990. Backpropagation is sensitive to initial conditions. *Complex Systems.* 4(3):269-280.

Kosko, B. 1990. Unsupervised learning in noise. *IEEE Transactions on Neural Networks.* 1(1):44-57.

Kozeny, J. 1927. Über kapillare leitung des wassers im boden. *Sitzungber. Akad. Wiss. Wien. Math. Naturw.* 136(Abt. IIa):271-306.

Krogh, A. and Vedelsby, J. 1995. Neural network ensembles, cross validation and active learning. In Tesauro, G., Touretzky, D.S. and Leen, T.K., editors, Advances in neural information processing systems. MIT Press, Cambridge, MA. 7:231-238.

Kumar, P., Merchant, S.N. and Desai, U.B. 2004. Improving performance in pulse radar detection using Bayesian regularization for neural network training. *Digital Signal Processing.* 14:438-448.

Leij, F. *et al.* 1996. The UNSODA unsaturated soil hydraulic database user's manual version 1.0. *National Risk Management Laboratory, Office of Research and Development, Cincinnati, OH.*

Lenhardt, R.J. 1984. Effects of clay-water interactions on water retention in porous media. PhD thesis. Oregon State University, Corvallis, OR, USA.

Levenberg, K. 1944. A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathematics II.* 2:164-168.

Levin, E., Tishby, N. and Solla, S.A. 1990. A statistical approach to learning and generalization in layered neural networks. *Proceedings ot the Institute of Electrical and Electronics Engineers.* 78(10):1568-1574.

Lilly, A. 1997. A description of the HYPRES database (hydraulic properties of European soils). In: Bruand, A., *et al.* eds. The use of pedotransfer functions in soil hydrology research. Proceedings second workshop of the project using existing soil data to derive hydraulic parameters for simulation modelling in environmental studies and in land use planning, Orléans, France, 10-12 October 1996. 161-184.

Low, P.F. 1961. Physical chemistry of clay-water interactions. *Advances in Agronomy.* 13:269-327.

Lu, B.L. and Ito, M. 1999. Task decomposition and module combination based on class relations: a modular neural network for pattern classification. *IEEE Trans. Neural Networks.* 10(5):1244-1256.

MacKay, D.J.C. 1992a. Bayesian interpolation. *Neural Computation.* 4(3):415-447.

MacKay, D.J.C. 1992b. Bayesian methods for adaptive models. PhD Thesis, California Institite of Technology, Pasadena, California.

Marquardt, D.W. 1963. An algorithm for least-squares estimation of non-linear parameters. *Journal of the Society of Industrial and Applied Mathematics II.* 2:431-441.

McCormack, C. 1997. Adaptation of learning rule parameters using a meta neural network. *Connection Science.* 9(1):123-136.

McCullagh, P. and Nelder, J.A. 1989. Generalized linear models, 2nd ed., Chapman & Hall, London.

McCulloch, W.S. and Pitts, W.H. 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*. 5:115-133.

McKenzie, N.J. and Jacquier, D.W. 1997. Improving the field estimation of saturated hydraulic conductivity in soil survey. *Australian Journal of Soil Research*. 35:803-825.

Millington, R.J. and Quirk, J.P. 1959. Permeability of porous media. *Nature*. 183:387-388.

Minasny, B. and McBratney, A.B. 2002. The neuro-m method for fitting neural network parametric pedotransfer functions. *Soil Science Society of America Journal*. 66:352-361.

Minasny, B., McBratney, A.B. and Bristow, K.L. 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. *Geoderma*. 93(3-4):225-253.

Minsky, M. 1969. In: Papert, S., editor. Perceptrons; an introduction to computational geometry. MIT Press, Cambridge, MA.

Mitchell, T.M. 1997. Machine Learning, Boston: WCB/McGraw Hill.

Moody, J.E. 1992. The effective number of parameters: an analysis of generalization and regularization in nonlinear learning systems, in Moody, J.E., Hanson, S.J., and Lippmann, R.P., *Advances in Neural Information Processing Systems*. 4:847-854.

Mualem, Y. 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*. 12:513-522.

Mualem, Y. and Dagan, F. 1978. Hydraulic conductivity of soils: Unified approach to the statistical models. *Soil Science Society of America Journal*. 42:392-395.

Mukkamala, S., Sung, A.H. and Abraham, A. 2004. Intrusion detection using an ensemble of intelligent paradigms. *Journal of Network and Computer Applications*. Article in Press.

Nemes, A., Schaap, M.G. and Wösten, J.H.M. 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. *Soil Science Society of America Journal*. 67:1093-1102.

Noakes, P.G. *et al.* 1993. 43k protein and acetylcholine receptors co-localize during the initial stages of neuromuscular synapse formation *in vivo*. *Developmental Biology*. 155:275-280.

Pachepsky, Ya.A. *et al.* 1982. Soil water retention as related to other soil physical properties. *Pochvovedenie*. 2:42-52.

Pachepsky, Ya.A., Timlin, D. and Várallyay, G. 1996. Artificial neural networks to estimate soil water retention from easily measurable data. *Soil Science Society of America Journal*. 60:727-773.

Patuwo, E., Hu, M.Y., and Hung, M.S. 1993. Two-group classification using neural networks. *Decision Sciences*. 24:825-845.

Paydar, Z. and Cresswell, H.P. 1996. Water retention in Australian soils, II. Prediction using particle size, bulk density and other properties. *Australian Journal of Soil Research.* 34:679-693.

Penny, W.D. and Roberts, S.J. 1997. Neural network predictions with error bars. Research report TR-97-1, Department of Electrical and Electronic Engineering, Imperial College, London.

Penny, W.D. and Roberts, S.J. 1998. Error bars for linear and nonlinear neural network regression models. Technical Report, Department of Electrical Engineering, Imperial College, London.

Perrone, M.P. and Cooper, L.N. 1993. When networks disagree: Ensemble method for neural networks, in Mammone, R.J., editor, Neural networks for speech and image processing. Chapman-Hall, London, 126-142.

Raats, P.A.C. 1992. A superclass of soils. In: Van Genuchten, M. T., Leij, F. J. and Lund, L. J. eds. Indirect methods for estimating the hydraulic properties of unsaturated soils. Proceedings of the international workshop on indirect methods for estimating the hydraulic properties of unsaturated soils, Riverside, California, 11-13 October 1989. 45-51.

Rajkai, K. and Várallyay, G. 1992. Estimating soil water retention from simpler properties by regression techniques. In: Van Genuchten, M. T., Leij, F. J. and Lund, L. J. eds. Indirect methods for estimating the hydraulic properties of unsaturated soils. Proceedings of the international workshop on indirect methods for estimating the hydraulic properties of unsaturated soils, Riverside, California, 11-13 October 1989. 417-426.

Raviv, Y. and Intrator, N. 1996. Bootstrapping with noise: an effective regularisation technique. *Connection Science* (special issue on combining estimators). 8:356-372.

Rawls, W.J., Brakensiek, D.L. and Saxton, K.E. 1982. Estimation of soil water properties. *Transactions of the American Society of Agricultural Engineers.* 25:1316-1320.

Rawls, W.J. *et al.* 2003. Effect of soil organic carbon on soil water retention. *Geoderma.* 116:61-76.

Richards, L.A. 1931. Capillary conduction of liquids through porous mediums. *Physics.* 1:318-333.

Ripley, B.D. 1994. Neural networks and related methods for classification. *Journal of the Royal Statistical Society, Series B, Methodological.* 56:409-456.

Roberts, S.J., Penny, W.D. and Pillot D. 1996. Novelty, confidence & errors in connectionist systems. Proceedings of IEE Colloquium on Intelligent Sensors and Fault Detection, September 1996, 1996/261: 10/1-10/6.

Rosenblatt, F. 1958. The Perceptron: A probabilistic model for information storage and organization in the brain. *Cornell Aeronautical Laboratory, Psychological Review.* 65(6):386-408.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. 1986. Learning representations by backpropagating errors. *Nature.* 323(9):533-536.

Schaap, M.G. and Bouten, W. 1996. Modelling water retention curves of sandy soils using neural networks. *Water Resources Research.* 32:3033-3040.

Schaap, M.G. and Leij, F.J. 1998a. Database-related accuracy and uncertainty of pedotransfer functions. *Soil Science.* 163:765-779.

Schaap, M.G. and Leij, F.J. 1998b. Using Neural Networks to predict soil water retention and soil hydraulic conductivity. *Soil and Tillage Research.* 47:37-42.

Schaap, M.G., Leij, F.L. and van Genuchten, M.Th. 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Science Society of America Journal.* 62:847-855.

Schaap, M.G., Leij, FJ. and van Genuchten, M.Th. 2001. ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology.* 251(3-4):163-176.

Scheinost, A.C., Sinowski, W. and Auerswald, K. 1997. Regionalization of soil water retention curves in a highly variable soilscape, i. Developing a new pedotransfer function. *Geoderma.* 78(3-4):129-143.

Sharkey, A.J.C. 1999. Combining artificial neural nets: ensemble and modular multi-net systems. Springer, London.

Sinowski, W., Scheinost, A.C. and Auerswald, K. 1997. Regionalization of soil water retention curves in highly variable soilscape, II. Comparison of regionalization procedures using a pedotransfer function. *Geoderma.* 78:145-159.

Snyder, V. 1996. Statistical hydraulic conductivity models and scaling of capillary phenomena in porous media. *Soil Science Society of America Journal.* 60:771-774.

Sollich, P. and Krogh, A. 1996. Learning with ensembles: When over-fitting can be useful. In Touretzky, D.S., Mozer, M.C. and Hasselmo, M.E. editors, *Advances in Neural Information Processing Systems.* 8:190-196. MIT Press. Cambridge, MA.

Stephens, M.A. 1974. EDF Statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association.* 69:730-737.

Tamari, S., Wösten, J.H.M. and Ruiz-Suárez, J.C. 1996. Testing an artificial neural network for predicting soil hydraulic conductivity. *Soil Science Society of America Journal.* 60:1732-1741.

Theng, B.K.G. 1973. The chemistry of clay-organic reactions, Wiley, New York.

Timlin, D.J., Ahuja, L.R. and Williams, R.D. 1996. Methods to estimate soil hydraulic parameters for regional scale applications of mechanistic models. In: Application of GIS to the modelling of non-point source pollutants in the vadose zone. *Soil Science Society of America Journal.* Special Publication 48:185-203.

Tomasella, J. and Hodnett, M.G. 1998. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. *Soil Science.* 163:190-202.

Tumer, K. and Ghosh, J. 1995. Order statistics combiners for neural classifiers. In: Proceedings of the world congress on neural networks. INNS Press, Washington, DC. I:31-34.

Tumer, K. and Ghosh, J. 1996. Error correlation and error reduction in ensemble classifiers. *Connection Science,* Special issue on combining artificial neural networks: ensemble approaches. 8(3&4):385-404.

U.N. Commission on Sustainable Development (UNCSD) 1997. Comprehensive assessment of the freshwater resources of the world. United Nations, New York.

USDA Natural Resource Conservation Service 1994. National Soil Pedon Database, Lincoln, NE.

van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal.* 44:892-898.

Veelenturf, L.P.J. 1995. Analysis and applications of artificial neural networks, New York: Prentice Hall.

Vereecken, H. *et al.* 1989. Estimating the soil moisture retention characteristic from texture, bulk density and carbon content. *Soil Science.* 148:389-403.

Wang Y. 1998. Active learning based on a hybrid neural network modeller. PhD Thesis, University of Abertay Dundee, Dundee. 42-45.

Warner, B., and Misra, M. 1996. Understanding neural networks as statistical tools. *The American Statistician.* 50:284-293.

Werbos, P.J. 1974. Beyond regression: new tools for prediction and analysis in the behavioral sciences. doctoral dissertation, Applied Mathematics, Harvard University, Boston, MA.

Widrow, B. and Hoff, M.E. 1960. Adaptive switching circuits. In 1960 IRE WESCON Convention Record, 4:96-104.

Weigand, S., Huberman, A., and Rumelhart, D.E. 1990. Predicting the future: a connectionist approach. *International Journal of Neural Systems.* 1(3):101-117.

Wolpert, D.H. 1990. A mathematical theory of generalization. *Complex Systems.* 4:151-200.

Wolpert, D.H. 1992. Stacked generalization. *Neural Networks.* 5:241-259.

World Health Organization (WHO) 1995. Community water supply and sanitation: needs, challenges, and health objectives. Report by the Director-General to the forty-eighth World Health Assembly, Geneva.

World Health Organization (WHO) 2004. [Online]. World water day 2004: water and disasters. The role of the World Health Organization. Available from the World Wide Web: URL: http://www.who.int/water_sanitation_health/wwd2004fs/en.

Wösten, J.H.M. and van Genuchten, M.Th. 1988. Using texture and other soil properties to predict the unsaturated soil hydraulic functions. *Soil Science Society of America Journal.* 52:1762-1770.

Wösten, J.H.M. *et al.* 1990. Functional sensitivity analysis of four methods to generate soil hydraulic functions. *Soil Science Society of America Journal.* 54:832-836.

Wösten, J.H.M. *et al.* 1998. Using existing soil data to derive hydraulic parameters for simulation models in environmental studies in land use planning. DLO-Staring Centre Rep.156. Wageningen, the Netherlands.

Wösten, J.H.M. *et al.* 1999. Development and use of a database of hydraulic properties of European soils. *Geoderma.* 90(3-4):169-185.

Wösten, J.H.M., Pachepsky, Ya.A. and Rawls, W.J. 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology.* 251:123-150.

Young, I.M., Crawford, J.W. and Rappoldt, C. 2001. New methods and models for characterising structural heterogeneity of soil. *Soil and Tillage Research.* 61:33-45.